VOLUME 39 SEPTEMBER 1953 NUMBER 9

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
OF THE
UNITED STATES OF AMERICA

EDITORIAL BOARD

LINUS PAULING, *Chairman* EDWIN B. WILSON, *Managing Editor*
ALEXANDER WETMORE, *Home Secretary* ROGER ADAMS, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS G. W. BEADLE GREGORY BREIT
R. E. CLELAND E. A. DOISY L. C. DUNN
W. W. RUBEY G. C. EVANS W. F. LIBBY
HARLOW SHAPLEY C. H. GRAHAM J. VON NEUMANN
F. E. TERMAN J. T. PATTERSON PAUL WEISS

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office and Home Office of the Academy
2101 Constitution Avenue, Washington 25, D. C.

Entered as Second-Class matter June 21, 1930, at the Post-Office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in section 1103, act of October 3, 1917. Authorized June 18, 1920.

Published Monthly Price per Annum, $7.50

OFFICERS OF THE ACADEMY

PRESIDENT ..... DETLEV W. BRONK
VICE PRESIDENT ..... EDWIN B. WILSON
FOREIGN SECRETARY ..... ROGER ADAMS
HOME SECRETARY ..... ALEXANDER WETMORE
TREASURER ..... WILLIAM J. ROBBINS

INFORMATION TO SUBSCRIBERS

Checks covering subscriptions to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES at the rate of $7.50 per annum should be made payable to the National Academy of Sciences, and sent to the Treasurer, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C. Single numbers $0.75.

Microfilm copies of the PROCEEDINGS are available to subscribers to the paper edition. The microfilm is a positive, furnished on metal reels, suitably labeled, and is distributed at the end of the volume year. Inquiries concerning purchase should be directed to University Microfilms, 313 N. First St., Ann Arbor, Michigan.

Past publications of the NATIONAL ACADEMY OF SCIENCES from 1863 to 1925 and of the NATIONAL RESEARCH COUNCIL from 1916 to 1925 are listed in the Index to the First Ten Volumes of these PROCEEDINGS. This Index appeared as Part 2 of Volume 13, No. 1, January, 1927, and is for sale separately at the price of $1.50. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Past volumes of the PROCEEDINGS may be obtained at five dollars per volume unless the sale of the volume would break a complete set of Volumes I to XXX. Single numbers may be obtained for fifty cents except where the sale of such numbers would break up a complete volume. Beginning with Volume 37 (1951), complete volumes are $7.50, and single numbers $0.75. Only forty complete sets of the PROCEEDINGS are available for sale —Volumes I to XXX price $200.00. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

THE NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL also issues technical publications reviewing and correlating recent research. The *Bulletin* and *Reprint and Circular Series* have been replaced by a single publication series, available individually, or by subscription at 25% discount. For a list of available publications, with prices and ordering procedures, address the Publications Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 39 September 15, 1953 Number 9

ALTERATION OF THE KINETIC PROPERTIES OF AN ENZYME BY THE BINDING OF BUFFER, INHIBITOR, OR SUBSTRATE

By Robert A. Alberty and Robert M. Bock

Chemistry Department and Biochemistry Department, University of Wisconsin

Communicated by Farrington Daniels, July 20, 1953

A number of the features of the catalytic properties of enzymes have been interpreted in terms of further elaborations of the simple mechanism of Michaelis and Menten.[1] Some treatments have shown that the kinetic constants for the over-all reaction cannot be interpreted in the simple way originally thought. Recently, Foster and Niemann[2] have discussed the theory for a difunctional enzyme combining with a difunctional substrate and mono- or difunctional inhibitor and have shown that the maximum initial velocity may be a measure of the equilibrium constants for various steps in the mechanism in addition to being a function of the rate constant for the breakdown of the enzyme-substrate complex. When the catalytic properties of enzymes are complicated by interaction with components of the buffer,[3–7] in addition to hydrogen ion, mechanisms may be enlarged to also take these effects into account. The mechanisms described in this paper have been developed to account for the following facts concerning the kinetics of fumarase which are not in accord with simple mechanisms used earlier. There is (1) a strong dependence of Michaelis constants and maximum initial velocities on buffer concentration and composition, figure 1 (*a*); (2) a variation of competitive inhibition constants with buffer concentration; (3) new types of inhibition, figure 1 (*b*) and (*c*); and (4) under certain conditions, markedly curved Lineweaver-Burk[8] plots, figure 1 (*d*). These effects appear to be related in that they are all accounted for by mechanisms which allow for the alteration of the properties of the enzyme by the binding of buffer, inhibitor, or even substrate in such a way that the enzymatic activity may be either enhanced or diminished. These effects are not of the nature of the ionic strength effects studied by Kistiakowsky and Shaw,[9] and such effects are neglected in the discussion of the following mechanisms.

Initial velocities for the reactions catalyzed by crystalline fumarase were obtained by recording the optical density changes at 210 to 280 mμ with a Beckman DUR spectrophotometer equipped to record optical density 0 to 0.097.[10] When fumarate was the substrate kinetic measure-

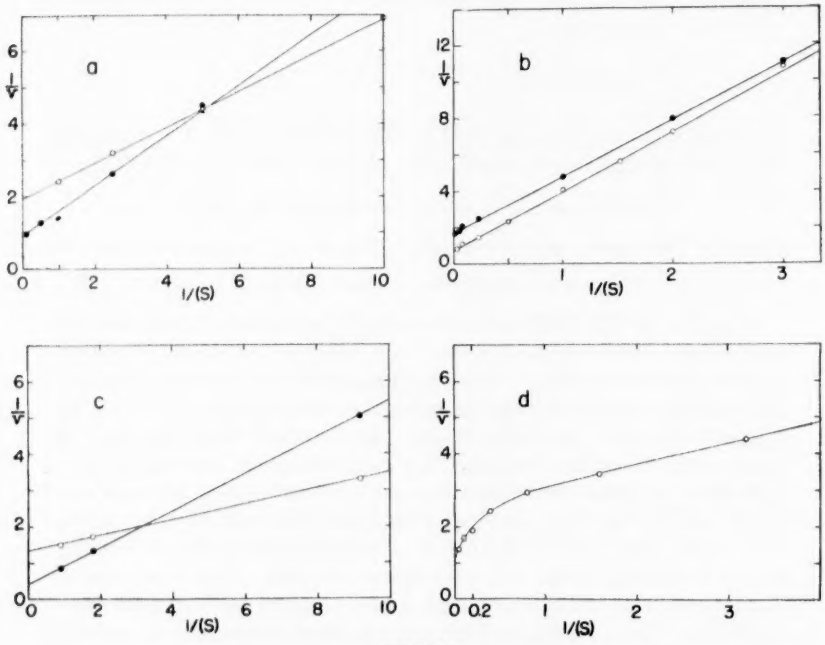


FIGURE 1

Reciprocal initial velocity versus reciprocal substrate concentration (millimolar). The velocities are in arbitrary units, and since the enzyme concentrations are not the same in (*a*), (*b*), (*c*), and (*d*), the maximum velocities may not be directly compared. (*a*) Fumarate, pH 7.3; ○, 0.015 *M* phosphate buffer; ●, 0.15 *M* phosphate buffer. (*b*) L-Malate, pH 7.5; ○, 0.05 *M* phosphate buffer; ●, 0.05 *M* phosphate buffer plus 0.1 *M* NaCl. (*c*) Fumarate, pH 8.07; ○, 0.005 *M* phosphate buffer; ●, 0.005 phosphate buffer plus 0.040 *M* succinate. (*d*) L-Malate, pH 7.4, in 0.05 ionic strength *tris*-(hydroxymethyl)-aminomethane-hydrochloric acid buffer.

ments were limited to the first 8% reaction, while only the first 2% reaction was used in the case of L-malate because of the unfavorable equilibrium.

Interaction with Buffer.—If, in addition to the basic reaction[1] between enzyme, E, and substrate, S, to yield product, P,

$$E + S \underset{}{\overset{K_{ES}}{\rightleftharpoons}} ES \xrightarrow{k_1} E + P, \qquad (I)$$

the buffer, B, can combine with the enzymatically active site

$$E + B \underset{}{\overset{K_{EB}}{\rightleftharpoons}} EB, \qquad (II)$$

or at another site at which it affects the kinetic constants,

$$E + B \underset{}{\overset{K_{BE}}{\rightleftharpoons}} BE, \qquad (III)$$

the following reactions may also be important:

$$BE + S \underset{}{\overset{K_{BES}}{\rightleftharpoons}} BES \xrightarrow{k_2} BE + P \qquad (IV)$$

$$BE + B \underset{}{\overset{K_{BEB}}{\rightleftharpoons}} BEB. \qquad (V)$$

The fact that BES is allowed to break down to yield product makes this mechanism analogous to the "non-total" inhibition discussed by Segal, Kachmar, and Boyer[11] and to the effects of "modifiers" discussed by Botts and Morales.[12] It will be assumed that in all subsequent derivations the equilibria for which the dissociation constants (K) are indicated are adjusted rapidly in comparison to the rate of appearance of product and that $(S) \gg (E)_0$, and $(B) \gg (E)_0$, where $(E)_0$ is the total molar concentration of the enzymatically active sites. It is not necessary to include steps such as $ES + B \rightleftharpoons BES$ since they are not independent of the above equilibria. For the mechanism represented by (I)–(V) the initial rate (v) of appearance of product at constant (B) will vary with substrate concentration according to the familiar expression

$$v = \frac{V'}{1 + K'_S/(S)}, \qquad (1)$$

where V', the maximum initial velocity, and K'_S, the Michaelis constant, are now given by

$$V' = k_1(E)_0 \frac{1 + k_2K_{ES}(B)/k_1K_{BE}K_{BES}}{1 + K_{ES}(B)/K_{BE}K_{BES}} \qquad (2)$$

$$K'_S = K_{ES} \frac{1 + (B)(1/K_{BE} + 1/K_{EB}) + (B)^2/K_{BE}K_{BEB}}{1 + K_{ES}(B)/K_{BE}K_{BES}}. \qquad (3)$$

Equations of this form may be used to represent the variation of V' and K'_S for fumarase such as that indicated in figure 1 (a). In sodium phosphate buffers of pH 7.3 relative values for V' for fumarate are 0.32, 0.47, 0.68, and 0.81 at 0.005, 0.015, 0.05, and 0.15 M phosphate, and the corresponding values of K'_S are 0.26, 0.38, 0.74, and 2.0 mM. It will be noted that according to this mechanism V' may be a function of buffer concentration even if $k_2 = 0$.

Interaction with Buffer and an Inhibitor.—If an inhibitor, I, which

combines only at the enzymatically active site is added, the mechanism may be represented by (I)–(VII).

$$E + I \underset{}{\overset{K_{EI}}{\rightleftharpoons}} EI \qquad (VI)$$

$$BE + I \underset{}{\overset{K_{BEI}}{\rightleftharpoons}} BEI \qquad (VII)$$

and the initial rate reaction rate is expressed by

$$v = \frac{V'}{1 + \frac{K'_S}{(S)}\left(1 + \frac{(I)}{K'_I}\right)}, \qquad (4)$$

where the competitive inhibition constant K'_I is

$$K'_I = K_{EI}\frac{1 + (B)(1/K_{EB} + 1/K_{BE}) + (B)^2/K_{BE}K_{BEB}}{(1 + K_{EI}(B)/K_{BE}K_{BEI})}. \qquad (5)$$

Thus K'_I is not simply the dissociation constant for an EI complex, as is commonly assumed. In the case of fumarase the variation of the competitive inhibition constant for sodium *trans*-aconitate (0.46, 0.97, and 1.50 mM at 0.005, 0.015, and 0.05 M sodium phosphate buffer of pH 7.5) may readily be expressed by Eq. (5).

If the inhibitor, like the buffer, can also combine with the enzyme so that the resulting complex still acts catalytically, but with different properties, the following two steps must be added (possible complexes IEB and IEI are ignored):

$$E + I \underset{}{\overset{K_{IE}}{\rightleftharpoons}} IE. \qquad (VIII)$$

$$IE + S \underset{}{\overset{K_{IES}}{\rightleftharpoons}} IES \overset{k_3}{\rightarrow} IE + P. \qquad (IX)$$

These are steps which would probably become important in the case of a poor inhibitor which is tested at a relatively high concentration. For the mechanism represented by (I)–(IX), the variation of v with (S) is represented by Eq. (1) with

$$V' = k_1(E)_0\frac{1 + k_2K_{ES}(B)/k_1K_{BE}K_{BES} + k_3K_{ES}(I)/k_1K_{IE}K_{IES}}{1 + K_{ES}(B)/K_{BE}K_{BES} + K_{ES}(I)/K_{IE}K_{IES}} \qquad (6)$$

$$K'_S = K_{ES}\frac{1 + (B)(1/K_{EB} + 1/K_{BE}) + (B)^2/K_{BE}K_{BEB} + (I)(1/K_{EI} + 1/K_{IE} + (B)/K_{BE}K_{BEI})}{1 + K_{ES}(B)/K_{BE}K_{BES} + K_{ES}(I)/K_{IE}K_{IES}}. \qquad (7)$$

Thus the intercept of the Lineweaver-Burk plot may be either larger or smaller than in the absence of inhibitor and the inhibition may not fall into any of the recognized classes.[13] The type of inhibition represented

by Eqs. (6) and (7) is a general type which results in linear Lineweaver-Burk plots and reduces to competitive, non-competitive, uncompetitive and apparent competitive[11] cases under special conditions. In the case of fumarase the inhibition by chloride at pH 7.3, figure 1 (*b*) and, in general, by poor inhibitors such as succinate, figure 1 (*c*), does not fit into the usual classes.

If steps (I)–(IX) apply and $k_3 = 0$, V may decrease with increasing (I). This is an effect which may also occur in the case of a difunctional enzyme, difunctional substrate, and monofunctional inhibitor as pointed out by Foster and Niemann.[2]

Interaction with Buffer and Substrate.—The explanation of the effects represented in figure 1 (*a*)–(*c*) in terms of alteration of catalytic properties of the enzyme by the binding of buffer or inhibitor suggests that substrate itself may have a similar effect. If this is the case, the following two steps are added to the mechanism represented by Eqs. (I)–(V):

$$\mathrm{E} + \mathrm{S} \underset{}{\overset{K_{\mathrm{SE}}}{\rightleftharpoons}} \mathrm{SE}. \qquad \text{(X)}$$

$$\mathrm{SE} + \mathrm{S} \overset{K_{\mathrm{SES}}}{\rightleftharpoons} \mathrm{SES} \overset{k_4}{\rightarrow} \mathrm{SE} + \mathrm{P}. \qquad \text{(XI)}$$

The variation of initial rate with substrate concentration is given by an equation of the type

$$v = \frac{a + b/(S)}{1 + c/(\mathrm{S}) + d/(\mathrm{S})^2} \qquad (8)$$

where a, b, c, and d are constants for a particular buffer concentration. Figure 1 (*d*) indicates that this type of equation satisfactorily represents the data for fumarase in a *tris*-(hydroxymethyl)-aminomethane chloride buffer of pH 7.5. Such activation by substrate is in contrast with the inhibition which is frequently observed at high substrate concentrations in the case of some other enzymes. As pointed out by Kistiakowsky and Rosenberg[3] this type of rate equation also results if there are two types of sites which are either different and independent or which are identical but interact in pairs. It may be very difficult to distinguish between these various possibilities. It is characteristic of Eq. (8) that straight line plots will be obtained by the Lineweaver and Burk method at very high or very low substrate concentrations, so that if experiments fall accidentally in one of these regions the data appears to conform to the simple Michaelis-Menten mechanism. This may be seen when (8) is written as

$$v = \frac{V_2 + V_1K_2/(1 - V_1/V_2)(\mathrm{S})}{1 + K_2/(1 - V_1/V_2)(\mathrm{S}) + K_1K_2/(1 - V_1/V_2)(\mathrm{S})^2} \qquad (9)$$

where
$$V_1 = k_1(\mathrm{E})_0 \frac{1 + k_2K_{ES}(\mathrm{B})/k_1K_{BE}K_{BES}}{1 + K_{ES}/K_{SE} + K_{ES}(\mathrm{B})/K_{BE}K_{BES}}$$

$$K_1 = K_{ES} \frac{1 + (\mathrm{B})(1/K_{BE} + 1/K_{EB}) + (\mathrm{B})^2/K_{BE}K_{BEB}}{1 + K_{ES}/K_{SE} + K_{ES}(\mathrm{B})/K_{EB}K_{BES}}$$

$$V_2 = k_4(\mathrm{E})_0$$

$$K_2 = K_{SE}K_{SES}[(1 - k_1/k_4)/K_{ES} + 1/K_{SE} + (\mathrm{B})(1 - k_2/k_4)/K_{BE}K_{BES}]$$

As (S) → 0, Eq. 9 becomes $v = V_1/(1 + K_1/(\mathrm{S}))$. It should be noted that V_1 and K_1 are the same as V' and K's given by Eqs. (2) and (3) except for the additional term K_{ES}/K_{SE} which appears in the denominator as a result of equilibrium (X).

As (S) → ∞, Eq. (9) becomes $v = V_2/(1 + K_2/(\mathrm{S}))$, and it is of interest to note that according to this mechanism the limiting velocity at high substrate concentration (V_2) is independent of the buffer concentration. The importance of this type of mechanism for the control of the rate of an enzymatic reaction in vivo is that the rate may be greatly increased in the presence of a considerable excess of the substrate for the enzyme.

Acknowledgment.—The authors are indebted to Dr. Vincent Massey and Mr. Carl Frieden for many helpful discussions. This investigation was aided by grants from the National Science Foundation and the Research Committee of the Graduate School of the University of Wisconsin.

[1] Michaelis, L., and Menten, M. L., *Biochem. Z.*, **49,** 333 (1913).
[2] Foster, R. J., and Niemann, C., these PROCEEDINGS, **39,** 371 (1953).
[3] Kistiakowsky, G. B., and Rosenberg, A. J., *J. Am. Chem. Soc.*, **74,** 5020 (1952).
[4] Fasman, G. D., and Niemann, C., *Ibid.*, **73,** 1646 (1951).
[5] Kistiakowsky, G. B., Mangelsdorf, P. C., Jr., Rosenberg, A. J., and Shaw, W. H. R., *Ibid.*, **74,** 5015 (1952).
[6] Chance, B., and Harting, J., *Federation Proc.*, **12,** 188 (1953).
[7] Massey, V., *Biochem. J.*, **53,** 57 (1952).
[8] Lineweaver, H., and Burk, D., *J. Am. Chem. Soc.*, **56,** 568 (1934).
[9] Kistiakowsky, G. B., and Shaw, W. H. R., *Ibid.*, **75,** 2751 (1953).
[10] Bock, R. M., and Alberty, R. A., *Ibid.*, **75,** 1921 (1953).
[11] Segal, H. L., Kachmar, J. F., and Boyer, P. D., *Enzymologia*, **15,** 187 (1952).
[12] Botts, J., and Morales, M., *Trans. Far. Soc.* **49,** 696 (1953).
[13] Wilson, P. W., in Lardy, H. A., *Respiratory Enzymes*, Burgess Publ. Co., Minneapolis, 1949, pp. 21-30.

A LIGHT ACTIVATION PHENOMENON IN THE ENZYMATIC AND NONENZYMATIC REDUCTION OF TETRAZOLIUM SALTS*

BY WALTER J. NICKERSON AND JOSEPH R. MERKEL†

DEPARTMENT OF MICROBIOLOGY, RUTGERS UNIVERSITY

Communicated by S. A. Waksman, July 9, 1953

Photoreduction of 2,3,5-triphenyltetrazolium chloride (TTC) has been reported by a number of investigators.[1-5] According to Gierlach and Krebs,[4] photoreduction of TTC occurs at wave-lengths shorter than 3650 Å. Anaerobic conditions have been reported to facilitate reduction of TTC in the dark.[3] We have found that the photoreduction of TTC solutions occurs more rapidly under alkaline conditions.

In studies on the mechanism of reduction of TTC by washed cell suspensions of yeasts,[6] it was noted that certain cultures did not reduce the tetrazolium in the dark. It was found that the ability to reduce TTC in the dark was related to the age of the cultures. Cells from cultures of yeasts incubated 12 to 14 hrs. in liquid media with continuous agitation reduced TTC equally well in the light or in the dark, whereas, cells from cultures incubated for 72 hrs. had lost most of their ability to reduce TTC in the dark (table 1). It was also noted that various inhibitors of TTC reduction were more effective in the dark than in the light. In particular, the inhibitory effect of monoiodoacetate and of zinc on 72-hr. cultures was pronounced in the dark and negligible in the light.

Although ultra-violet radiations can, under certain conditions, cause reduction of TTC, it can be seen from the table that the microbial reduction described in the present studies is an enzymatic reaction. From studies with inhibitors, it appears that hydrogen donating systems within the cell are affected by the photoactivation. Attempts to activate the TTC solutions and the yeast suspensions by irradiation, before they were mixed, were unsuccessful.

Brodie and Gots[7] have reported that tetrazolium is reduced by a flavoprotein enzyme. It therefore appears as if the light activation is directed at this type of enzyme system. It also appears that this system has been altered by aging the yeast cultures. Evidence for the involvement of a flavin-type substance in the light-activated reaction has been obtained.

The properties of a non-enzymatic light-activated system for the reduction of tetrazolium salts have been studied (table 2). The system consisted of the tetrazolium salt (6×10^{-3} *M*), riboflavin or riboflavin phosphate (10^{-4} *M*), cysteine·HCl (2.5×10^{-3} *M*) and pH adjusted within the range 7.3–8.5. The photoreductions were carried out using 10 ml. volumes in uncovered 94-mm. petri dishes. A 375-watt Sylvania super

flood light was used for the illumination at a distance of 30–50 cm., with a $CuSO_4$ solution ($M/100$) of 7 mm. thickness acting as a heat filter.

Both TTC and blue tetrazolium[5] (BT) were studied in the system and it was found that under optimum conditions, TTC was more easily reduced than BT. Earlier work[6] had indicated that increased tetrazol-reductase activity could be obtained by adding various metal-chelating agents to resting cell suspensions of yeasts. It was therefore of interest to find that the disodium salt of ethylenediaminetetraacetic acid (Na_2-

TABLE 1

PHOTOACTIVATION OF ENZYMATIC TETRAZOLIUM REDUCTION BY WASHED RESTING CELLS ON *Candida albicans*

	TETRAZOLIUM REDUCTION AFTER 2-HOUR INCUBATION					
	12-HOUR CULTURES				72-HOUR CULTURES	
	PARENT STRAIN		FILAMENTOUS MUTANT		FILAMENTOUS MUTANT	
SYSTEM WITH TTC	LIGHT	DARK	LIGHT	DARK	LIGHT	DARK
Boiled cells	None	None	None	None	None	None
Boiled cells + NaCN (0.005M)	None	None	None	None	None	None
Washed cells	Trace	Trace	+++	+++	+++	Trace
Washed cells + Zn^{++} (0.005M)	Trace	Trace	Trace	Trace	++	None
Washed cells + Cu^{++} (0.005M)	None	None	None	None	None	None
Washed cells + NaCN (0.005M)	++++	++++	++++	++++	++++	++
Washed cells + Na_2EDTA	++++	+++	...	...	++++	++
Washed cells + iodo-acetate	...	...	...	...	+++	Trace

The yeasts were grown in 100 ml. volumes of a 1 per cent peptone and 0.5 per cent glucose medium at 28°C. with continuous agitation. The cells were harvested by centrifugation and washed twice with distilled water. Suspensions with the appropriate concentration of substrate or inhibitor were made up to 5 ml. in Pyrex test tubes. One set of tubes was exposed to indirect, diffused sunlight and the other set was placed in a dark cabinet. After two hours the tubes were examined for reduction of TTC (original concentration = 0.02 per cent) and scored on a comparative bases, i.e., trace = faint reduction, + = visible reduction, etc.

EDTA) could replace cysteine in the photoactivated system for the reduction of BT. In the case of TTC it was found that the addition of Na_2EDTA alone, at a pH on the alkaline side, permitted the photochemical reduction to occur. Riboflavin was not necessary for the photoreduction of TTC but greatly stimulated the reduction.

These findings led to the obvious conclusion that a metal ion(s) was associated with the tetrazolium salts and stabilizes them against photoreduction. Spectrographic analyses of samples of BT and TTC revealed the presence of large amounts of metals in BT, whereas TTC was found

to be relatively free of metals (table 3). One of the reaction components for the synthesis of BT is obtained commercially as a stabilized powder with 5 per cent zinc chloride and 20 per cent aluminum sulfate.[8]

TABLE 2

RIBOFLAVIN SENSITIZED NON-ENZYMATIC PHOTOCHEMICAL REDUCTION OF TETRAZOLIUM DYES

SYSTEM	REDUCTION AFTER 5% MINUTES ILLUMINATION
BT	None
BT + glucose + glycine	None
BT + cysteine (pH 8.8)	None
BT + R + cysteine (pH 8.0)	++++
BT + R + cysteine (pH 6.9)	None
BT + RP (pH 9.0)	None
BT + RP (pH 7.4)	None
BT + RP + Na_2EDTA (pH 7.3)	++++
BT + RP + cysteine (pH 7.8)	+

R = riboflavin.
RP = riboflavin phosphate.

TABLE 3

SPECTROGRAPHIC ANALYSIS OF METAL CONTAMINATION OF TETRAZOLIUM DYES, RIBOFLAVIN, AND DISODIUM VERSENE

METAL	BT	TTC	RIBOFLAVIN	RIBOFLAVIN PHOSPHATE	Na_2EDTA	BLANK
Al	Strong	Weak	Weak	Strong	Medium	...
Ca	Strong	...	Weak	Strong	...	...
Cr	Medium	...	...	...	...	...
Cu	Weak	Trace	Weak	Weak	Trace	...
Fe	Weak	Weak	Very weak	Weak	Trace	Trace
Mg	Strong	Strong	Strong	Very strong	Strong	Trace
Mn	Weak	...	...	...	Trace	...
Na	Strong	...	...	Very strong	Very strong	...
Zn	Very strong	...	...	...	...	...

BT—from Dajac Laboratories
TTC—an Eastman product
Na_2EDTA—Bersworth Chemical Company
Riboflavin and riboflavin phosphate—Hoffman-La Roche

The data presented account for the role of a metal chelating agent in the photochemical reduction, and for the greater sensitivity of TTC to photoreduction.[9]

Riboflavin appears to function in the photochemical reaction as the activated substance which reduces the tetrazolium compound. Although it has an unusual complexing affinity for iron,[10] it does not function merely

as another metal chelating agent. A previously irradiated solution of riboflavin phosphate to which a small amount of Na_2EDTA had been added would reduce TTC in the dark, indicating a relatively long life for the "active riboflavin."

It is interesting to compare our findings with those of Galston[11] who reported a riboflavin-sensitized *photo-oxidation* of indoleacetic acid and related compounds under acidic conditions in which light activated riboflavin was presumed to act as a hydrogen carrier between the substrate and its oxidized product. Our results seem to be analogous, but under the conditions we used the reaction is shifted to the reductive rather than the oxidative side, possibly by the shift in hydrogen-ion concentration. Riboflavin in the present studies seems to be functioning in a manner which has been proposed for chlorophyll:[12]

$$(Me^{2+}C)\cdot H_2O \xrightarrow{h\nu} (Me^{2+}C)^*\cdot H_2O$$
$$(Me^{2+}C)^*\cdot H_2O \rightarrow (Me^{+}C)\cdot H_2O^{+}$$
$$(Me^{+}C) + Ox \rightarrow (Me^{2+}C) + Ox^{-}$$

where $Me^{2+}C$ = metal-chlorophyll complex.

Future experiments should indicate the nature of the "active riboflavin" molecule in this type of photoreduction.

These reactions, interesting in themselves, have several implications. They emphasize the necessity of taking the possible occurrence of photochemical reactions into account when studying enzymatic reduction of tetrazolium compounds, or when studying flavin-catalyzed reductions in general. We have noted that many of our cytochemical studies with TTC have been complicated by photochemical reduction of the tetrazol in wet mounts under illumination from a microscope lamp. Secondly, the effect of Na_2EDTA focuses attention on the importance of metal chelation in maintaining oxidation-reduction balances in intracellular systems.

Acknowledgment.—The authors wish to thank Mr. Niel Shimp and Prof. Arthur Prince of the Soils Department of the Agricultural Experiment Station for making the spectrographic analyses recorded in table 3.

* Journal Series Paper—New Jersey Agricultural Experiment Station, Rutgers University, the State University of New Jersey, Department of Microbiology. This investigation was supported in part by a grant from the National Institute of Health, Public Health Service.

† Waksman-Merck Postdoctoral Fellow in the Natural Sciences.

[1] Ried, W., *Angewandte Chem.*, **64**, 391 (1952).

[2] Bielig, H. J., Kausche, G. A., and Haardick, H., *Z. Naturforsch.*, **4b**, 80 (1949).

[3] Weygard, F., and Frank, I., *Z. Naturforsch.*, **3b**, 377 (1948).

[4] Gierlach, Z. S., and Krebs, A. T., *Am. J. Roentgenol. Radium Therapy*, **62**, 559 (1949).

[5] Hausser, I., Jerchel, D., and Kuhn, R., *Chem. Ber.*, **82**, 195–199 (1949).
[6] Merkel, J. R., and Nickerson, W. J., (in manuscript).
[7] Brodie, A. F., and Gots, J. S., *Arch. Biochem. Biophys.*, **39**, 165 (1952).
[8] Rutenburg, A. M., Gofstein, R., and Seligman, A. M., *Cancer Research*, **10**, 113–121 (1950).
[9] Dyar, M. T., *Am. J. Botany*, **40**, 20 (1953).
[10] Albert, A., *Biochem. J.*, **47**, XXVII (1950).
[11] Galston, A. W., PROC. NATL. ACAD. SCI., **35**, 10–17 (1949).
[12] Weiss, J., Symposia of the *Soc. for Exptl. Biol.*, No. V, pp. 144–145 (1951).

*THE INTERACTION OF AUXIN AND LIGHT IN THE GROWTH RESPONSES OF PLANTS**

BY JAMES L. LIVERMAN AND JAMES BONNER

KERCKHOFF LABORATORIES OF BIOLOGY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 17, 1953

Introduction.—Many of the growth and developmental responses of higher plants are influenced by light or by the interplay of light and dark. The control of the morphogenetic aspects of plant behavior appears to be exerted through mechanisms which are distinct from and in some cases at least independent of the photosynthetic machinery. Light affects directly, for example, the growth of hypocotyls and mesocotyls, coleoptile elongation, leaf expansion, the germination of certain seeds, and the pathway of bud differentiation. The photochemical control of floral initiation, the phenomenon of photoperiodism, constitutes a particularly well described instance of the morphogenetic role of light. Despite our recognition of light as a dominant factor in plant development, we have however little more than descriptive knowledge of the processes involved.

A second factor which we recognize today as of general significance in the control of plant morphogenesis is the chemical control exerted by hormones of the auxin group. Auxins control many morphogenetic responses, and it so happens in addition, that many of these auxin-controlled responses are further affected by light. That this is so is particularly evident in the much-studied cases of coleoptile growth and of floral induction. In both of these instances light and exogenous auxin supplied in appropriate dosages act in the same direction, promoting the growth of coleoptiles[1] and the flowering of long day plants[2] or on the other hand suppressing the flowering of short day plants.[3] Added light and added auxin can, to a considerable extent, replace one another in the control of these two quite different plant responses.

We may logically inquire whether there may exist some relation between the effects of light and those of auxin. We might ask whether light in some

way brings about its effects through influences upon the auxin system of the plant itself. It has already been shown by Galston[4] that appropriate light treatment increases the destruction of auxin by plant tissue through activation of the indoleacetic acid oxidase system. This paper will be concerned only with those cases in which growth is promoted by exposure to light. Red light, which we know to promote the elongation of oat coleoptiles, appears to exert its influence by inducing the production in plant tissue of an auxin receptor which thereby becomes able to combine with auxin to form the growth active auxin-receptor complex. It is this complex which appears to be responsible for many and perhaps for all of the auxin regulated plant responses.[5] Light in the near infra-red wave-length region inhibits the auxin-induced growth of oat coleoptiles and appears to do so by destruction of the active auxin-receptor complex. The products of the infra-red mediated reaction include an inactive form of the receptor, a form not capable of recombining with auxin. Finally it will be shown that the facts available concerning certain of the auxin-light mediated plant responses are compatible with and may be interpreted in terms of the concept of a cyclical light controlled process. In the morphogenetic photocycle red light promotes the formation of auxin-receptor from an inactive auxin-non-receptive precursor. Infra-red light on the other hand promotes the destruction of the auxin-receptor complex with regeneration of auxin-non-receptive precursor.

Experimental Approach.—That the auxin molecule is active in eliciting the growth response only after it has become combined to an appropriate receptor entity within the cell is evident from the fact that auxin-induced growth may be accurately described in terms of classical enzyme kinetics.[6] We may represent the growth process as controlled by the reactions of Eq. (1) below.

$$E + S \underset{}{\overset{K_s}{\rightleftharpoons}} ES \xrightarrow{k} \text{Growth} \qquad (1)$$

E of Eq. (1) represents the auxin receptive moiety within the plant while S represents the auxin molecule. These two combine to form the growth-active complex, ES. The system is further described by the parameters K_s and k which are, respectively, the disassociation constant of ES and the rate constant which relates ES to final growth rate. Since in many systems neither the absolute total concentration of E nor the absolute value of k is readily determinable, their product may be replaced by the parameter V_{max}, the growth rate under conditions of non-limiting S concentration in which all E is converted to ES.

There are many ways in which light could conceivably effect an auxin mediated growth system formulated as in Eq. (1). Thus light might influence either the formation or the inactivation of E, of S, or of ES. Alternatively, light might affect the combination of E with S. And, finally,

light might influence only the constant k which relates the concentration of ES to over-all growth rate. In order to come to a decision as to which alternative, if any, is the correct one it is desirable to use a test system in which the several variables may be controlled. Such a system has been found in excised sections of the oat coleoptile. Auxin is the limiting factor in the growth of coleoptile sections under the usual experimental conditions. This fact permits experimental control of the concentration of S and thus of the concentration of ES within the tissue.

Materials and Methods.—All of the experimental work reported here is based on measurement of the initial rate of elongation of excised sections

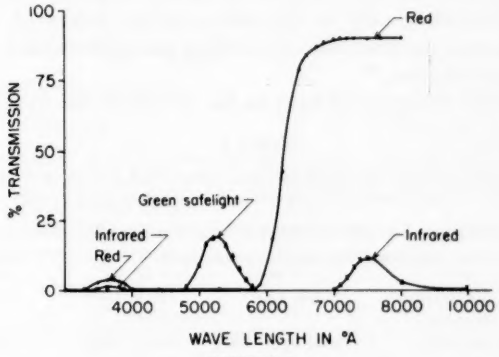


FIGURE 1

Spectral qualities of filters used with varied light sources for illumination of plants. *Green source* (used for preparative procedures): light from 40-watt Mazda filtered through Corning Glass Filter (CG) No. 3385, Color Specification (CS) 3-71 and CG No. 4010, CS 4-64. *Red source:* Light from 2 daylight fluorescent tubes filtered through red cellulose acetate film. Spectral quality of filter determined with Beckman spectrophotometer. *Infrared source:* Light from 6-200 watt photofloods filtered through 8-mm. CG No. 5874, CS 7-39 and red cellulose acetate film.

of Avena coleoptiles floated in solutions containing indoleacetic acid (IAA). The culture of the Avena seedlings was as described by Foster, *et al.*[5] When the seedlings had attained an age of 72–84 hours, depending upon the particular experiment, they were selected for uniformity. For the majority of experiments, coleoptiles ranging from 1.0 to 1.5 cm. in length were used. In certain experiments, as noted below, older coleoptiles with lengths of 2.0–2.5 cm. or of 2.5–3.0 cm. were used. A single section 3.8 or 5 mm. in length was cut from the apical end of each coleoptile 2–3 mm. below the tip. The sections were pooled and randomly distributed in lots of 20 to individual Petri dishes containing 20 ml. of the test solution. The basal medium contained 0.005 M potassium maleate buffer (pH 4.5) and 3 per

cent sucrose. The auxin (indoleacetic acid, IAA) adjusted to pH 4.5, was added to the basal medium as indicated below.

The sections were allowed to grow for 6 or 16 hours at a temperature of 25–26°C. and their length then measured.

The spectral qualities of the various filters used in illumination of the plants are shown in figure 1.

The Response to Red Light.—That the growth of the Avena coleoptile is influenced by red light has been known since the work of Vogt, Lange, and others.[6] During the first 80 hours, growth of the seedling Avena coleoptile is negligible in total darkness[1,7] but proceeds rapidly in low intensities of red light.[1] The action spectrum for the promotion of Avena coleoptile growth[6] corresponds closely to the action spectra found for the photoperiodic response,[8] the promotion of seedling leaf growth,[9] and the promotion of seed germination.[10]

The promotive effect of red light on the growth of the Avena coleoptile

TABLE 1

THE EFFECT OF RED LIGHT IN ACCELERATING GROWTH RATE OF AVENA COLEOPTILE SECTIONS IN A NON-LIMITING CONCENTRATION OF AUXIN (6×10^{-6} *M* IAA)

Initial length of sections 3.8 mm. Corrected for growth in auxin-free medium. A-51.

GROUP	SEQUENCE AND NATURE OF TREATMENT			GROWTH IN MM./ SECTION/16 HR.
	30 MIN.	30 MIN.	15 HR.	
1	Red light, basal medium	Dark, IAA	Dark, IAA	2.13 ± 0.10^a
2	Dark, IAA	Red light, IAA	Dark, IAA	2.25 ± 0.13
3	Dark, basal medium	Dark, IAA	Dark, IAA	1.70 ± 0.12

[a] Difference from Group 2 not significant. Difference from Group 3 significant at 1% level.

cannot be mediated by effects on auxin production since red light increased growth rate of excised sections which are supplied with non-limiting amounts of auxin.[1] It can further be shown that red light elicits its growth promotive effect even when auxin is not present in the tissue at the time of exposure to light. For these experiments sections from dark grown plants were cut under low-intensity green light. The sections were pooled and divided into three lots which were then treated separately. Group 1 was exposed directly to red light for 30 minutes, returned to darkness, and IAA added. Group 2 was supplied with IAA and allowed to remain in the dark for 30 minutes before being moved to red light. After a 30 minute exposure to red light the plants were returned to the dark. Group 3 received 30 minutes' treatment in auxin-free basal solution in the dark after which the auxin was added and the sections maintained in darkness. All groups were measured after 16 hours' incubation. The results of a typical experiment are presented in table 1. It is clear that a brief red light treat-

ment promotes subsequent elongation of coleoptile sections whether or not auxin is present during the exposure to light.

Further information concerning the manner in which red light interacts

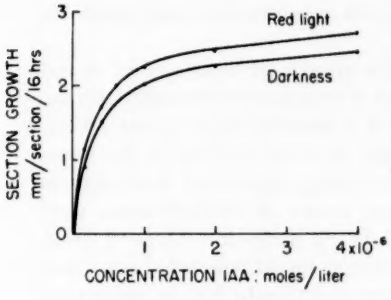


FIGURE 2

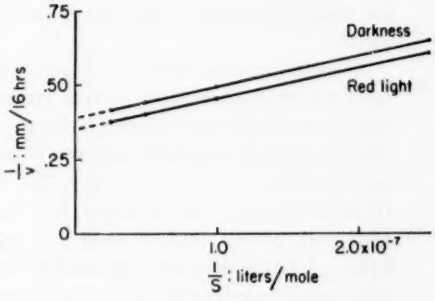


FIGURE 3

Fig. 2. Growth rate of Avena coleoptile sections as a function of indoleacetic acid (S) concentration. Red-light-treated sections received 30 minutes of illumination at beginning of 16-hr. incubation period.

Fig. 3. Data of Fig. 2 plotted in double reciprocal form; 1/growth rate as a function of 1/IAA concentration. Red light increases V_{max} but is without effect on K_s.

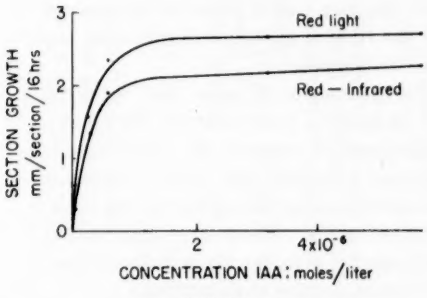


FIGURE 4

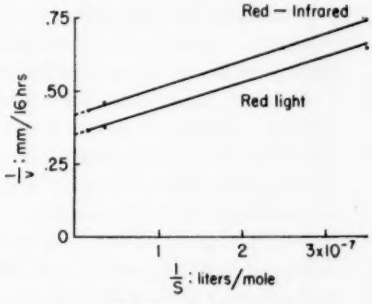


FIGURE 5

Fig. 4. Growth of Avena coleoptile sections as a function of indoleacetic acid (S) concentration. All sections pretreated with red light and then subjected to 45 minutes of infra-red light at the beginning of 16-hr. incubation period.

Fig. 5. Data of Fig. 4 plotted in double reciprocal form; 1/growth rate as a function of 1/IAA concentration. Infra-red light decreases V_{max} as does an uncompetitive inhibitor.

with the coleoptile to increase response to auxin is furnished by the data of figure 2 in which growth velocity is plotted as a function of concentration of auxin supplied to the sections. Figure 3 presents the same data in reciprocal form following the Lineweaver-Burk treatment. It is clear from

figure 3 that red light serves to increase V_{max} but leaves K_s uninfluenced. The action of red light is then to be interpreted as due either to an increase in E_{total} (increase in total number of auxin-receptive entities) or as due to an increase in k. Red light does not alter the affinity of the receptive entity for auxin nor does it alter the quantitative requirement of the tissue for auxin.

The Response to Infra-Red Light.—The growth promotive effect of red light on Avena coleoptile section growth is reversed by infra-red light, the maximum effectiveness being attained at a wave-length of about 7300Å. This reversal is obtained only when auxin has been supplied to the tissue before infra-red light is given. The following experiment demonstrates these relations. Plants which had been grown in total darkness were sectioned, pooled, and divided into 3 lots, all under low-intensity green light. All lots were next exposed to red light for 30 minutes. Lot 1 was transferred to infra-red light for 45 minutes and then to dark at which time auxin was added. Lots 2 and 3 were moved from red light to darkness at which time they received auxin. The sections of lot 2 were incubated with auxin for 1 hour in the dark and then given 45 minutes of infra-red light followed by darkness during the remainder of the experiment. All sections were measured after 6 or 16 hours' incubation in auxin solution. The results of three such experiments are shown in table 2.

The data show that infra-red light is effective only if given in the presence of auxin. If auxin is not present at the time of exposure to infra-red, the light is ineffective in causing inhibition of elongation.

Further information concerning the interaction of auxin and infra-red light is contained in figures 4 and 5. In figure 4 growth rate is plotted as a function of auxin concentration while figure 5 presents the same data in double reciprocal form. It is clear that infra-red light acts to decrease V_{max} but leaves K_s unchanged. This effect is manifest only when the infra-red light is given in the presence of added auxin. These facts suggest that infra-red light may attack and decompose the complex ES but that it does not influence E, S, or the k relating ES concentration to growth rate.

It is apparent from the data of figures 2 and 4 and from the data of tables 1 and 2 that the effects of red and of infra-red light in the growth of Avena coleoptile sections are only quantitative ones. Under the conditions of the experiments and with the plant material used red light increases growth rate by some 25% or more while infra-red inhibits by the same amount. The coleoptile evidently possesses the ability to carry out both the red and infra-red mediated reactions at reduced rates in the dark. Thus sections cut in the dark from dark grown plants show no further inhibition of growth when given infra-red light. Only after dark adapted plants have been supplied with red light does infra-red light become effective in reducing growth. It appears therefore, that infra-red light and darkness act in the same direc-

TABLE 2

INHIBITION OF RED LIGHT INDUCED GROWTH OF AVENA COLEOPTILE SECTIONS BY INFRA-RED LIGHT AND THE RELATION OF THIS INHIBITION TO THE PRESENCE OF AUXIN (2.86×10^{-6} *M* IAA)

Growth rates corrected for growth in auxin-free medium and calculated on basis of 5-mm. sections. Original lengths of coleoptiles for experiments 1, 2, and 3 were 1–1.5, 1.5–2.0, and 2.2–2.8 cm., respectively.

GROUP	SEQUENCE AND NATURE OF TREATMENT				GROWTH IN MM./SECTION/6 HR. EXPT. 1	EXPT. 2	EXPT. 3
1	Red light, basal medium, 30 minutes	Infra-red light, basal medium, 45 minutes	Dark, IAA	Dark, IAA	2.33 ± 0.06[a]	2.02 ± 0.08[b]	1.85 ± 0.06[c]
2	Red light, basal medium, 30 minutes	Dark, IAA, 1 hour	Infra-red light, IAA, 45 minutes	Dark, IAA	2.09 ± 0.05	1.55 ± 0.15	1.49 ± 0.12
3	Red light, basal medium, 30 minutes	Dark, IAA, 1 hour	Dark, IAA, 45 minutes	Dark, IAA	2.51 ± 0.04	2.10 ± 0.1	1.75 ± 0.07

[a] Difference from Group 3 not significant. Difference from Group 2 significant at 2% level.
[b] Difference from Group 3 not significant. Difference from Group 2 significant at 1% level.
[c] Difference from Group 3 not significant. Difference from Group 2 significant at 1% level.

TABLE 3

INTERACTION OF RED AND INFRA-RED LIGHT IN GROWTH OF AVENA COLEOPTILE SECTIONS IN A NON-LIMITING AUXIN CONCENTRATION (3×10^{-6} *M* IAA)

Corrected for growth in auxin-free medium. Sections maintained in auxin solution during entire experimental period. Original length of sections, 5 mm. A-66.

GROUP	SEQUENCE AND NATURE OF TREATMENT 30 MIN.	30 MIN.	30 MIN.	30 MIN.	4 HR.	GROWTH IN MM./SECTION/6 HR.
1	Dark	Red light	Dark	Dark	Dark	2.51 ± 0.07
2	Dark	Red light	Infra-red light	Dark	Dark	1.96 ± 0.09[a]
3	Dark	Red light	Infra-red light	Red light	Dark	2.54 ± 0.05
4	Dark	Infra-red light	Dark	Dark	Dark	1.76 ± 0.08
5	Dark	Dark	Dark	Dark	Dark	1.67 ± 0.07

[a] Difference from Groups 1 and 3 significant at 1% level. Difference from Group 4 not significant. Difference from Group 5 significant at 3% level.

tion. It is also evident that although growth rate of dark adapted plants is increased by red light there is not an absolute requirement for such light. It has already been pointed out that dependence of coleoptile growth on red light decreases as the seedling grows. Since sections can be cut only from coleoptiles which have grown, the experimental material for this work has necessarily been restricted to coleoptiles which have partially lost their red light requirement.

It has been suggested above that infra-red light depresses Avena section growth rate by removal of ES. This transformation appears to involve the production of a form of E which is not itself able to combine with S but which may be transformed into S-receptive E through the action of red light. This is shown by the following experiment which demonstrates that the effect of infra-red light is immediately and completely reversed by subsequent exposure to red light.

Sections were cut from dark grown plants and divided into 5 lots. All lots were pretreated for 30 minutes in the dark with 3×10^{-6} M IAA. Lots 1, 2, and 3 next received 30 minutes of red light. Lot 1 was then removed directly to darkness. Lots 2 and 3 were given 30 minutes of infra-red light after which lot 2 was removed to darkness while lot 3 was given 30 minutes of red light and then moved to darkness. Lot 4 received only infra-red light for 30 minutes before being moved to darkness. Lot 5 was placed directly in darkness without light treatment. The plants were grown for a total of 6 hours in auxin solution. The results shown in table 3, indicate that growth rate in this system is almost wholly determined by the quality of the last light treatment given. Furthermore it may now be concluded that the effect of red light is specifically to increase E_{total} (E + ES). This follows from the fact of red light reversal of infra-red inhibition, an inhibition already shown to result from decrease in E_{total}.

The Photocycle.—The experimental facts brought forth above can be formulated as in reactions 2, 3, and 4 below.

$$E_{inactive} \xrightarrow[\text{(dark slowly)}]{\text{Red light}} E \qquad (2)$$

$$E + S \rightleftharpoons ES \qquad (3)$$

$$ES \xrightarrow[\text{(dark slowly)}]{\text{Infrared}} E'_{inactive} + S' \qquad (4)$$

According to Eq. (2), red light causes the rapid formation of E from an auxin-non-receptive precursor. The E thus produced combines with auxin to form auxin-receptor complex (reaction 3), a reaction which is not influenced by light so far as can be detected (no effect of light on K_s). Auxin-receptor complex is, however, rapidly degraded to an inactive form by infra-red light (reaction 4). We have already seen however that the growth-inactive product of infra-red action may be again transformed to an active

form by red light. It appears possible therefore that $E'_{inactive}$ of reaction (4) may be interchangeable or identical with $E_{inactive}$ of reaction (2) and both might be termed $E_{precursor}$ or E_p. If this is assumed to be so, then reactions (2) to (4) may be simplified and combined in cyclic form as indicated in figure 6. The cyclic formulation of the reactions which have been proposed as relating light to auxin action will be referred to as the photocycle.

Interpretation of Plant Responses in Terms of the Photocycle.—The principles of the cyclic interconversion of E_p, E and ES discussed above are of general interest since it appears that many of the otherwise unrelated descriptive facts concerning the morphogenetic effects of light and auxin can be comprehended within this one conceptual framework. This is true for example of photoperiodism. The flowering of short day plants is induced by the production within the leaf of a flowering hormone. The formation of this hormone takes place only when the leaf is exposed to a period or periods of continuous darkness which exceed a definite critical length. The effectiveness of a dark period longer than the critical may be nullified by interruption with red light[8, 11] or by application of auxin to the leaf.[3] The application of an auxin antagonist to the leaf promotes flowering even when the dark periods are shorter than the critical. Applied auxin antagonists also antidote the effect of red light interruption of the dark period[12] as well as the effect of applied auxin.[3] These facts suggest that the induction of flowering may be a response to a drop in effective auxin level within the leaf, a drop which is somehow achieved in dark periods longer than the critical. They suggest further that red light acts to interfere with the drop in effective auxin level which normally occurs during darkness. These changes in effective auxin level do not appear to be clearly reflected in any changes in the level of free, uncombined auxin in the leaf.

THE MORPHOGENETIC PHOTOCYCLE

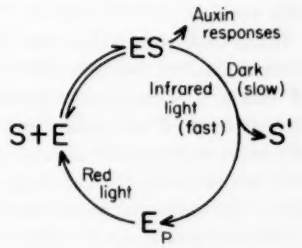


FIGURE 6

The morphogenetic photocycle. A hypothesis concerning the interaction of auxin and light in plant tissues. *S*, represents the auxin molecule; *E*, the auxin receptive entity; *ES*, the auxin-receptor complex; E_p, the auxin-non-receptive precursor of *E*.

The photoperiodic behavior of short day plants is interpretable in terms of the photocycle if we assume that the flowering response is governed by the level of ES within the leaf. During the day the leaf is exposed to light and all E_p is converted to E. This combines with auxin to form ES. During the dark period ES decays to E_p. If the night is sufficiently long, the level

of ES drops below that critical for the initiation of flowering. If, however, the night is interrupted by a flash of red light, E_p is reconverted to E and the supply of ES is hence regenerated. ES can at any time be converted to E_p by infra-red light which is capable of reversing the effect of red light interruption or, in general, of shortening the critical dark length for short day plants.[13] The fact that leaves exhibit a variety of responses to added auxin indicates that S is normally limiting and that leaves contain free E which is converted to ES only in the presence of added auxin. When a short day plant is treated with auxin at the beginning of the dark period the initial level of ES is increased and the dark period needed for decay of ES to the value critical for flowering is thereby lengthened.[14]

With some but not with all short day plants, leaves kept sufficiently long in darkness lose the ability to respond immediately to light unless auxin is also supplied.[14] This implies that the free auxin level of the leaf may also drop in darkness, an implication which has not, however, been experimentally verified.

It has been shown that the short day plant, Kalanchoe, will not flower in continuous darkness but will flower if given a single flash of red light each 24 hours.[16] This seemingly strange behavior is understandable in the present terms if we make the further assumption that flowering in a short day plant requires not merely a level of ES below some critical but rather a level within some critical range. In constant darkness ES is continuously converted to E_p and would ultimately be totally depleted were it not for the reconversion of E_p to E by the intermittent flashes of light.

Floral induction in long day plants takes place in response to day lengths in excess of and night lengths less than an appropriate critical. Plants maintained on a regime of nights longer than the critical however can be caused to flower by either supplemental red light[15] or, application of auxin to the leaves at the beginning of each dark period.[2] The behavior of long day plants suggests, in the present terms, that floral induction occurs only when the level of ES is maintained above some threshold value throughout the dark period. The kinetics of the formation of E from E_p in red light and of the decay of ES in darkness would appear to be generally similar to those in short day plants. The flowering response of long day plants to ES level is, however, the reverse of the response of short day plants as normally studied. It may be noted that the response of long day plants to ES level is paralleled by the behavior of the short day Kalanchoe which flowers in darkness only if given intermittent light.

Leaf Expansion.—The leaves of many seedling plants do not expand unless they are exposed to light. The energies required are low and the action spectrum, with a maximum at 6600Å, is generally similar to that of other light controlled morphogenetic responses.[9] A particularly convenient system with which to work is the seedling bean leaf,[17] excised disks

of which expand in response to red light. The response to red light is magnified in the presence of added auxin and is completely reversed by infra-red light.[18] These relations are similar to those found in the oat coleoptile. Their interpretation in the present terms is as follows: young bean leaves require ES for expansion. They can synthesize some S (auxin) but the conversion of E_p to E is almost totally red light dependent. The decay of ES is either rapid (in infra-red light) or slow (in dark).

Seed Germination.—Certain seeds, as those of lettuce, require light in order to germinate. The action spectrum for the promotion of lettuce seed germination is similar to those for other morphogenetic responses.[10] The effect of red light in promoting germination may be immediately negated by infra-red light and decays also, although more slowly, in darkness.[10] In order to determine whether these observations fit into the present scheme it is necessary first to know whether ES is essential to seed germination, i.e., whether seed germination is an auxin-mediated process. Added auxins are, in general without effect on germination at low concentrations and are either ineffective or inhibitory at higher concentrations.[19] It is well known, however, that seeds normally contain auxin[20] and that varied changes in auxin content occur during germination.[21] Germinating seedlings also show a variety of responses to added antiauxins and to auxins in high concentrations. It would appear then that if germination is indeed an auxin mediated process, seeds normally possess both E and S. Lettuce seed would, in the present terms, be remarkable in containing E_p rather than E. The red light requirement would be based on the requirement for the conversion of E_p to E preparatory to the formation of ES which would then lead to germination.

Conclusion.—The information presented above suggests that many heretofore unrelated processes which are affected by both light and auxin, including red light induced growth of the Avena coleoptile, the photoperiodic response, leaf expansion, and seed germination, may all have their basis in a common mechanism. All of these processes have approximately the same action spectra with effective peaks at *ca.* 6600Å, all are freely reversed by infra-red light, and all appear to be auxin-dependent. It has been shown that in the case of the Avena coleoptile, red light appears to exert its effect through the generation of an auxin-receptive entity, E, within the plant and that infra-red light acts to decompose the active complex ES into an auxin-non-receptive entity. In the coleoptile these processes also proceed thermally although at reduced rates. This is particularly true of the infra-red mediated decomposition of ES. The present paper calls attention to the fact that a concept of the cyclic interconversion of auxin-non-receptive precursor, auxin receptor, and receptor-auxin complex by light and dark can be applied without violence to the interpretation of these varied plant responses. The concept presented, that of the photo-

cycle, at once suggests a variety of experimental approaches which may be used to further test its validity in each individual instance.

* Report of work supported in part by the National Science Foundation and in part by the Lederle Laboratories Division, American Cyanamid Co.

[1] Schneider, C. L., *Am. J. Bot.*, **28,** 878 (1941). Went, F. W., and Thimann, K. V., *Phytohormones*, Macmillan Co., New York, 1937.

[2] Liverman, J., and Lang, A., Abst. Ann. Meeting Am. Soc. Plant Physiol., Cornell, 1952, p. 36.

[3] Thurlow, J., and Bonner, J., *Am. J. Bot.*, **34,** 603 (1947). Bonner, J., and Thurlow, J., *Bot. Gaz.*, **110,** 613 (1949).

[4] Galston, A. W., and Baker, R., *Am. J. Bot.* (1953); (in press).

[5] Foster, R. J., McRae, D. H., Bonner, J., PROC. NATL. ACAD. SCI., **38,** 1014 (1952).

[6] Vogt, E., *Z. Bot.*, **7,** 193 (1915). Johnston, E. S., *Smith. Misc. Coll.*, **96,** 1 (1937). Lange, S., *Jahrb. Wiss. Bot.*, **71,** 1 (1929).

[7] Weintraub, R. L., and Price, L., *Smith. Misc. Coll.*, **106,** 1 (1947).

[8] Parker, M. W., Hendricks, S. B., Borthwick, H. A., and Scully, N. J., *Bot. Gaz.*, **108,** 1 (1946).

[9] Parker, M. W., Hendricks, S. B., Borthwick, H. A., and Went, F. W., *Am. J. Bot.* **36,** 194 (1949).

[10] Borthwick, H. A., Hendricks, S. B., Parker, M. W., Toole, E. H., and Toole, V. K., PROC. NATL. ACAD. SCI., **38,** 662 (1952). Flint, L. H., and McAlister, E. D., *Smith. Misc. Publ.*, **94,** No. 5 (1935).

[11] Hamner, K. C., and Bonner, J., *Bot. Gaz.*, **100,** 388 (1938).

[12] Bonner, J., *Ibid.*, **110,** 625 (1949). Liverman, J., Diss., Calif. Inst. Tech., 1952.

[13] Borthwick, H. A., Hendricks, S. B., and Parker, M. W., PROC. NATL. ACAD. SCI., **38,** 929 (1952).

[14] Salisbury, F., and Liverman, J., Abst. Ann. Meeting Am. Soc. Plant Physiol., Western Sect., Santa Barbara, 1953.

[15] Withrow, R. B., and Richman, M. W., *Illum. Eng. Soc. Trans.*, **29,** 65 (1934). Katunskij, V. M., *Compt. rend. acad. sci. U.R.S.S.*, **3,** 303 (1936).

[16] Harder, R., and Gummer, G., *Planta*, **35,** 88 (1947).

[17] Miller, C. O., *Arch. Biochem. Biophys.*, **32,** 216 (1951).

[18] Liverman, J., and Starr, L.; (unpublished work, 1953).

[19] Mayer, A., and Evenari, M., *Bull. Research Council Israel*, **1,** 125 (1951).

[20] Haagen-Smit, A. J., Dandliker, W., Wittwer, S., and Murneek, A. E., *Am. J. Bot.*, **33,** 118 (1946).

[21] Cholodny, N., *Planta*, **23,** 289 (1935).

THE CONDUCTANCE OF SOLUTIONS OF SALTS IN ORTHODICHLOROBENZENE

BY FILIPPO ACCASCINA, E. LOWELL SWARTS, PHILIP L. MERCIER, AND CHARLES A. KRAUS

METCALF RESEARCH LABORATORY, BROWN UNIVERSITY

Communicated July 3, 1953

Although the dielectric constant of ethylene chloride differs little from that of ethylidene chloride, 10.23 and 10.00, respectively, the dissociation constants of salts dissolved in these two solvents differ markedly. Thus, according to Healey and Martell,[1] the dissociation constants of tetraethylammonium picrate, tetra-*n*-butylammonium picrate and tetraethylammonium nitrate in ethylene chloride are, respectively, 4 5, 5.0, and 9.9 times those in ethylidene chloride at 25°. A similar result was found by Ramsey and Colichmann[2] for *o*-chlorophenyltrimethylammonium perchlorate. Since the dielectric constants of these two solvents have practically the same value, the observed differences of dissociation constants must be ascribed to constitutional factors relating to the solvent molecules.

It seemed of interest to investigate this subject further by measuring the conductance of the same salts in another solvent having approximately the same dielectric constant as those already studied. For this purpose we have employed orthodichlorobenzene whose dielectric constant at 23.4° is 9.927.[3]

Conductance measurements were carried out with a Jones Bridge using an alternating current in the audio-frequency range. The conductance cells employed were of the Erlenmeyer type[4] and solutions were prepared by the dilution method developed in this laboratory.[5]

The orthodichlorobenzene as supplied by the producers[6] was of excellent quality and had a specific conductance of 1×10^{-9}; no attempt was made to separate possible isomers. It was refluxed over activated alumina and distilled under reduced pressure. The distillate was stored over powdered, activated alumina and solvent was withdrawn into the conductance cell through a fritted filter as needed. The density of the solvent at 25° was found to be 1.300 g./ml. and the viscosity 0.01271 poise. At 22°, the index of refraction for the *D* line of sodium was 1.550 in good agreement with values given in the literature. The dielectric constant[8] was found to be 9.927 at 23.4°; Smyth[7] reports 9.93 at 25°. The boiling range was 178–179° at atmospheric pressure.

Salts were prepared according to conventional methods and purified by recrystallization from suitable solvents as follows: tetraethylammonium picrate (m.-p. 260–261°), tetra-*n*-butylammonium picrate (m.-p. 91°),

and tetra-*n*-amylammonium picrate (m.-p. 75°) were recrystallized from 95% ethanol; tetra-*n*-propylammonium picrate (m.-p. 116.5°) was recrystallized from methanol; tetrabutylammonium nitrate (m.-p. 121°) from benzene and octadecyltributylammonium iodide (m.-p. 101°) and octadecyltrimethylammonium picrate (m.-p. 131–133°) from acetone. Tetrabutylammonium iodide (m.-p. 145–147°) was recrystallized from benzene. With the exception of octadecyltrimethylammonium picrate, two or more series of measurements were carried out with each salt, which was recrystallized between series. Conductance values agreed within the

TABLE 1

CONDUCTANCE OF SALTS IN *o*-DICHLOROBENZENE AT 25°

$C \times 10^4$	Λ_c	$C \times 10^4$	Λ_c	$C \times 10^4$	Λ_c
Et_4NPi		Pr_4NPi		Bu_4NPi	
2.653	9.978	2.984	8.953	4.358	7.702
1.551	12.36	1.556	11.55	2.927	9.028
0.9005	15.22	0.8006	14.81	1.603	11.38
0.5230	18.57	0.4191	18.54	1.158	12.82
0.2916	22.63	0.2185	22.71	0.6354	15.85
0.1606	27.13	...	...	0.3744	18.82
...	...	...	...	0.2303	21.71
Am_4NPi		Octd Me_3NPi		Octd Bu_3NI	
5.094	7.015	2.110	3.471	5.244	4.159
2.702	8.991	1.588	3.896	3.040	5.228
1.898	10.30	0.8785	5.076	2.108	6.107
1.120	12.52	0.4457	6.837	1.183	7.767
0.5002	16.50	0.2164	9.301	0.7138	9.544
0.3163	19.01	0.1298	11.44	0.4330	11.62
0.1697	22.54	...	...	0.2215	12.31
Bu_4NI		Bu_4NNO_3			
5.845	4.513	6.871	3.828		
3.102	5.860	3.370	5.155		
1.775	7.415	1.254	7.854		
1.043	9.235	0.5293	11.23		
0.4900	12.50	0.2064	16.23		
0.3152	14.79	...	...		
0.2063	17.22	...	...		

limit of experimental error (0.1%). All measurements were carried out at 25 ± 0.01° in an oil thermostat.

The results of conductance measurements will be found recorded in table 1 where concentrations are given in column 1 and equivalent conductances, Λ, in column 2. Results are given for only one series in each case.

The data were treated by the method of Fuoss,[8] values of the functions $F(z)/\Lambda$ and $C\Lambda f^2/F(z)$ being plotted as ordinates and abscissas, respectively. Examples for two salts are shown in figure 1. It will be noted that the

plots are linear and that there is excellent agreement between the results of different series. The intercepts of the plots on the axis of ordinates yield the values of Λ_0 and the slopes yield values of the dissociation constants K of the ion-ion pair equilibrium. The weaker the salt, the greater is the extrapolation and the greater the uncertainty in the value of Λ_0. The value of K, however, can be determined quite accurately. In this connection, it may be noted that while conductance values may be determined with good relative precision, absolute values are somewhat uncertain due to uncertainty as to the purity of the salts. The only

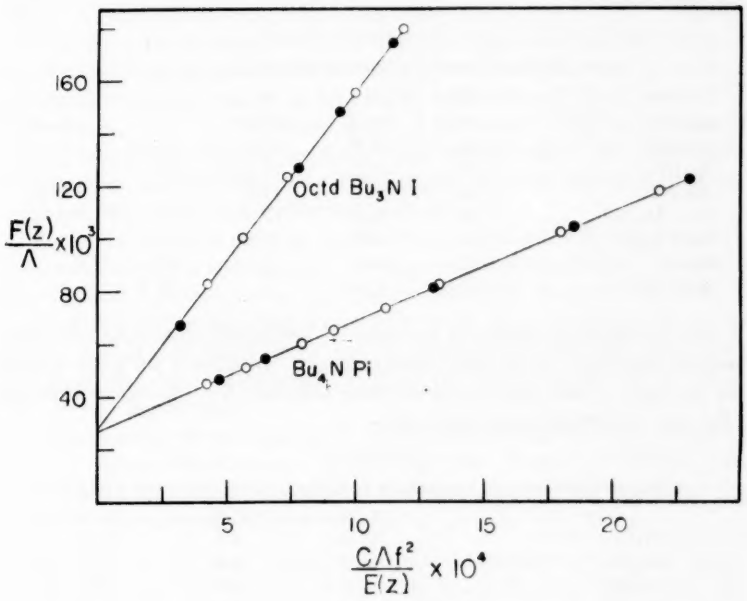


FIGURE 1

criterion of purity that we have is reproducibility of the conductance values of different samples of salt. While this usually suffices, it is not necessarily conclusive, particularly with salts whose ions contain long hydrocarbon chains.

In table 2 are collected values of derived constants for the different salts investigated. Values of the dissociation constant, K, are given in column 2 and of Λ_0, the limiting conductance, in column 3. Values of ion conductances are given for positive and negative ions in columns 4 and 5 respectively. In computing ion conductances, we have employed the method of Fowler and Kraus.[9] We have assumed that the conductance-

viscosity product, $\Lambda_0\eta$, for the tetrabutylammonium ion is the same in orthodichlorobenzene as it is in ethylene chloride. Observations with a considerable number of different solvents indicate that this product is constant within about 5%. In the absence of other means of evaluating ion conductances, this method serves as a useful approximation. In column 6 are given values of the reciprocal of the conductance-viscosity product, $1/\Lambda_0\eta$. This quantity is nothing more than the equivalent resistance of the electrolyte corrected for the viscosity of the solvent. Such correction makes it possible to compare the resistance of ions in different solvents.

TABLE 2

CONSTANTS FOR SALTS IN ORTHODICHLOROBENZENE AT 25°

SALT	$k \times 10^6$	Λ_0	Λ_0^+	Λ_0^-	$1/\Lambda_0^+\eta$
Et_4NP_i	1.14	47.4	26.3	...	2.99
Pr_4NP_i	1.32	41.7	20.6	...	3.82
Bu_4NP_i	1.71	37.3	16.2	21.1	4.85
Am_4NP_i	1.81	35.3	14.2	...	5.53
Octd Me_3NP_i	0.19	34.8	13.7	...	5.74
Octd Bu_3NI	0.63	35.0	11.9	...	6.61
Bu_4NI	0.64	39.3	...	23.1	...
Bu_4NNO_3	0.46	42.0	...	25.8	...

The dissociation constants for salts in orthodichlorobenzene are even smaller than they are in ethylidene chloride. In table 3 are given values for the ratio of the constants in ethylene chloride (K_1), ethylidene chloride (K_2) and orthodichlorobenzene (K_3).

TABLE 3

RATIO OF DISSOCIATION CONSTANTS IN $C_2H_4Cl_2$, CH_3CHCl_2, AND $C_6H_4Cl_2$

SALT	K_1/K_3	K_1/K_2	K_2/K_3
Et_4NP_i	13.9	4.6	3.05
Pr_4NP_i	14.7	4.9	3.01
Bu_4NP_i	13.2	5.0	2.65
Am_4NP_i	13.1	4.8	2.73
Octd Me_3NP_i	26.6	5.8	4.42
Et_4NNO_3	...	9.9	...
Bu_4NNO_3	25.6	...	...

For the picrates of symmetrical quaternary ammonium ions, from tetrabutyl to tetra-amyl, inclusive, the dissociation constant is approximately 13 times as great in ethylene chloride as in orthodichlorobenzene and it is approximately 4.8 times as great in ethylene chloride as in ethylidene chloride. The constant is approximately 2.75 times as great in ethylidene chloride as in orthodichlorobenzene and the ratio is somewhat greater for smaller ions.

Octadecyltrimethylammonium picrate is a very weak salt in ortho-

dichlorobenzene and in this solvent the K_1/K_3 ratio is 26.0. The K_1/K_2 ratio is also high for this salt being 5.8. For tetrabutylammonium nitrate, the K_1/K_3 ratio is 25.6 and for tetraethylammonium nitrate the K_1/K_2 ratio is 9.9. The dissociation constant of salts is particularly dependent on the nature of the negative ion.

The dielectric constant of orthodichlorobenzene is only 10% less than that of ethylene chloride and 7% less than that of ethylidene chloride. On the basis of Bjerrum's theory, we should not expect a large change in the dissociation constants of the picrates between ethylene chloride and orthodichlorobenzene. For tetra*iso*amylammonium nitrate in dioxane-water mixtures, the constant increases from 1×10^{-4} to 9×10^{-4} in going from a dielectric constant of 8.5 to 11.9.[10] It has been pointed out that the higher constant for salts in ethylene chloride may be due to rotation of the dipoles with respect to each other under the action of the field of adjacent ions.[1, 11] On the other hand, in the case of ethylidene chloride and orthodichlorobenzene, we have fixed dipoles while the dissociation constants of salts in these two solvents differ by larger amounts than one might expect from their dielectric constants.

There appears to be some ground for believing that the energy necessary to separate a pair of ions in a given medium may not be capable of complete interpretation in terms of a macroscopic dielectric constant. Indeed, it is somewhat surprising that this energy is so well accounted for on so simple a basis. It seems likely that the dissociation constants of ion pairs are dependent on constitutional factors of the solvent molecules as well as of the ions themselves.

Considerable interest attaches to the conductance of the three negative ions in orthodichlorobenzene, particularly the picrate and iodide ions. Below are given values of the ratio of the conductance of the iodide and the picrate ions in several different solvents.

Solvent:	$C_6H_4Cl_2$	$C_2H_4Cl_2$	C_5H_5N	$C_6H_5NO_2$	$(CH_3)_2CO$	H_2O
$\Lambda_0I^-/\Lambda_0Pi^-$:	1.05	0.96	1.44	1.28	1.31	2.53

Since the picrate ion contains 16 atoms other than hydrogen, we should expect the conductance of the iodide ion to be much greater than that of the picrate ion. Actually, it is only 5% greater in orthodichlorobenzene. From the above table, it will be noted that in ethylene chloride the conductance of the iodide ion is 4% less than that of the picrate ion. For some reason, the iodide ion has an exceptionally low conductance in the two halogenated hydrocarbon solvents. For the other solvents listed, the ratio varies from 1.28 for nitrobenzene to 2.53 for water.

Considering that the picrate ion contains 16 atoms other than hydrogen, its conductance is high when compared with that of quaternary ammonium ions. Thus, in orthodichlorobenzene, the conductance of the

picrate ion is 21.1 while that of the tetrabutylammonium ion (17 atoms) is 16.2 and that of the tetrapropylammonium ion (13 atoms) is 20.6. It may be noted that the conductance of the nitrate ion with 4 atoms is 17% greater than that of the picrate ion and 11% greater than that of the iodide ion. While our knowledge of the effect of constitutional factors on the conductance of negative ions is limited, it seems clear that their conductance is not primarily determined by the number and size of their constituent atoms.

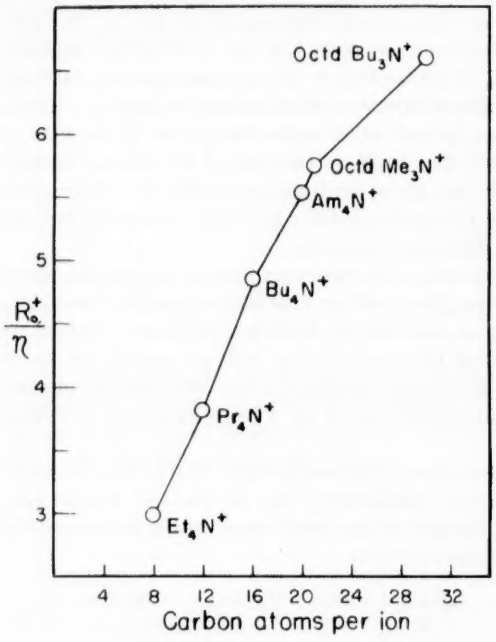


FIGURE 2

While we cannot change the size and structure of negative ions in any systematic manner, we can do this conveniently with the quaternary ammonium ions. However, in examining the conductance of such ions as a function of the number of carbon atoms that they contain, it is preferable to employ the reciprocal of ion conductances rather than this conductance directly. When ion conductances are plotted as a function of the number of carbon atoms that they contain, a highly flexed curve is obtained. In contrast, when the reciprocals of conductance are plotted against the number of carbon atoms, the curve approximates linearity, particularly for larger ions. Obviously, we have the relation

$$1/\Lambda = KC \times 10^{-3} = R_e$$

where K is the specific conductance of the solution, C is its concentration in equivalents per liter, and R_e is the equivalent resistance. As C decreases indefinitely, R_e approaches a limiting value, R_0. In comparing the properties of ions in different solvents, it is advantageous to employ resistance-viscosity ratios R_0/η.

In figure 2 is shown a plot of values R_0^+/η for the quaternary ammonium ions versus the number of carbon atoms in these ions. From tetraethyl to tetraamylammonium, inclusive, the plot is very nearly linear. The maximum deviation of any experimental point is 4%. The octadecyltrimethylammonium ion seems to lie on this same line, while the resistance of the octadecyltributylammonium ion lies markedly lower. A similar behavior has been found in other solvents.[12]

In orthodichlorobenzene, the equation of the straight line is:

$$R_0^+/\eta = 2.99 + n \times 0.211$$

where n is the number of carbon atoms in the ion. Plots of R_0^+ for the same ions (Et_4N^+–Am_4N^+) in other solvents are also approximately linear. The slope for orthodichlorobenzene is greater than for other organic solvents but markedly lower than it is for water. For water, orthodichlorobenzene, ethylidene chloride, and ethylene chloride, the coefficient n has the values: 0.250, 0.211, 0.198, and 0.177, respectively.

Ions that are smaller than the tetraethylammonium ion have a resistance which is relatively much greater than that of larger ions. For example, in acetone, the resistance of the ammonium ion is greater than that of the tetramethylammonium ion and is nearly equal to that of the tetraethylammonium ion.[12] In water, however, the resistance of ions increases regularly and by approximately equal increments in going from the ammonium to the tetra-*n*-amylammonium ion.[13] The deviation of the experimental values of R_0^+/η for water differs from the equation

$$R_0^+/\eta = 1.52 + n \times 0.250$$

by not more than 5%. Comparing the resistance of ions in water with that in other solvents, small ions have a relatively low resistance in water while large ions have a relatively high resistance. In non-aqueous solvents, small ions meet with resistance that appears to be quite unrelated to their size. If we are to account for the motion of ions in a viscous medium, it will probably be necessary to take into account the interaction of the charges on the ion with the dipoles of the solvent molecules. At the same time, it will probably be necessary to take into account structural factors of ions as well as of solvent molecules.

[1] Healey, F. H., and Martell, A. E., *J. Am. Chem. Soc.*, **73**, 3296 (1951).

[2] Ramsey, J. B., and Colichmann, E. L., *Ibid.*, **69**, 3041 (1947).

[3] We are indebted to Dr. Robert H. Cole of this laboratory for supplying us with this datum.

[4] Kraus, C. A., and Fuoss, R. M., *J. Am. Chem. Soc.*, **55**, 21 (1933).

[5] Mead, D. J., Fuoss, R. M., and Kraus, C. A., *Trans. Faraday Soc.*, **32**, 594 (1936).

[6] We are indebted to the Dow Chemical Company of Midland, Mich., for supplying us with this material.

[7] Smyth, C. P., Morgan, S. O., and Boyce, J. C., *J. Am. Chem. Soc.*, **50**, 1536 (1928).

[8] Fuoss, R. M., *Ibid.*, **57**, 1488 (1935).
[9] Fowler, D. L., and Kraus, C. A., *Ibid.*, **62**, 2237 (1940).
[10] Fuoss, R. M., Mead, D. J., and Kraus, C. A., *Ibid.*, **55**, 1014 (1933).
[11] Denison, J. T., and Ramsey, J. B., *J. Chem. Phys.*, **18**, 170 (1950).
[12] McDowell, M. J., and Kraus, C. A., *J. Am. Chem. Soc.*, **73**, 3293 (1951).
[13] Daggett, Jr., H. M., Bair, E. J., and Kraus, C. A., *Ibid.*, **73**, 799 (1951).

---

*THE EFFECT OF INDOLE-3-ACETIC ACID ON DATURA EMBRYOS**

BY J. RIETSEMA, S. SATINA, AND A. F. BLAKESLEE

GENETICS EXPERIMENT STATION, SMITH COLLEGE, NORTHAMPTON, MASS.

Communicated July 13, 1953

Introduction.—In the literature on the culture of immature and mature plant embryos slight mention has been made of the effect of auxins. This is remarkable since the presence of considerable amounts of auxin in seeds, especially in the endosperm, has been repeatedly demonstrated.[1, 4, 5, 6, 8] Moreover, the data published on the effect of auxin on embryo growth are incomplete and do not always agree.

The beneficial effect of low concentrations of tryptophane on orchid embryos has been explained by Spoerl[14] as being indirect on the assumption that this amino acid is converted into indole-3-acetic acid (IAA). Curtis and Nichol[3] on the other hand found a definite inhibition by 2 p.p.m. IAA on orchid embryos during the early stages of development after germination of the seeds.

Pohl[10] described experiments showing that pure IAA exerts an inhibiting effect on germinating Avena seeds, but mentioned a growth stimulating effect of auxin-*a* obtained from human urine or endosperm. Voss[19] showed that IAA has an inhibiting effect on mature corn embryos in concentrations from 5 p.p.m. and up, rather than a growth-promoting influence.

The effect of auxin on conifer embryos has been studied by Sterling,[15] and Loo and Wang.[7] The former found that very young larch embryos grown on media with 10^{-2} p.p.m. IAA show enlarged and vacuolated cells, whereas the latter observed cell divisions in young Pinus and Keteleeria embryos caused by 10 p.p.m. IAA.

Tukey[16] treated embryos of peach, apricot, and other fruits with various growth substances but he did not obtain a consistent response at the concentrations used. Rijven[12] mentioned inhibition of embryos of *Capsella bursa-pastoris* at 1 and 10 p.p.m. IAA and a slight stimulation at 10^{-3} p.p.m.

Van Overbeek[17] suggested that IAA is responsible for a callus-like

growth and for the root inhibition of immature Datura embryos grown on media containing coconut milk. After extraction of the coconut milk with ether, the root-inhibiting properties disappeared. Siu[13] described experiments with Datura embryos grown on media with IAA. He observed a stimulation by 10^{-2} p.p.m. only in the presence of an active coconut milk extract. Without this extract, inhibition and callus growth were found.

It seems from this review that IAA in high concentrations inhibits growth of embryos and in low concentrations causes some stimulation. This is of particular interest for a better understanding of the mechanism of embryo abortion in incompatible crosses since very little is known about natural inhibitors of embryo growth. The object of this paper therefore, has been to obtain more detailed data on the effect of IAA on embryo growth in vitro.

Methods and Material.—Test material was embryos of *Datura stramonium* (Line 1, standard line), obtained from capsules developed after self-pollination.

The culture methods and the media were the same as used in previous studies.[11] Indole-3-acetic acid (Eastman Kodak, Rochester, N. Y.) was added to the medium in the desired quantities. Length measurements of the embryos were taken after 9 days incubation in the dark at 25–26°C.

The development of the embryos in the ovules depends on both external and internal factors, such as temperature and humidity in the greenhouse, season of the year, condition of the plant and location of the capsule on the plant. As these factors could not be kept constant throughout the course of this investigation, we encountered difficulties in obtaining embryos of the same stage for the successive experiments. To avoid this difficulty we classified the embryos according to their size into three classes, viz., embryos smaller than 0.6 mm. (young and late hearts[11]); those between 0.6 and 2 mm. (young torpedoes[11]); and those longer than 2 mm. (advanced torpedoes to nearly mature embryos[11]). The cotyledons, hypocotyls (inclusive of the radicle) and roots if present were measured individually; their average length was expressed in percentages of the control. In addition, attention was given to histological responses since the length in normal embryos usually corresponds to a definite stage in histological differentiation.

Results.—The elongation of the cotyledons proved to be largely independent of the auxin concentration. Concentrations lower than 10^{-2} p.p.m. did not affect the length and higher concentrations did so to a limited extent. Most sensitive are the cotyledons of the smallest embryos. Their growth with 1 p.p.m. is approx. 40 per cent of the control. For cotyledons of embryos in the torpedo stage or of nearly mature embryos these values are 65 and 85 per cent, respectively. With 10 p.p.m. the

growth values of the cotyledons are 8, 34, and 55 per cent for the three size classes. Figure 1 shows the relation between IAA concentration and the cotyledon growth. Each point is an average of 5–8 experiments.

The hypocotyls are slightly more susceptible to IAA. As figure 2

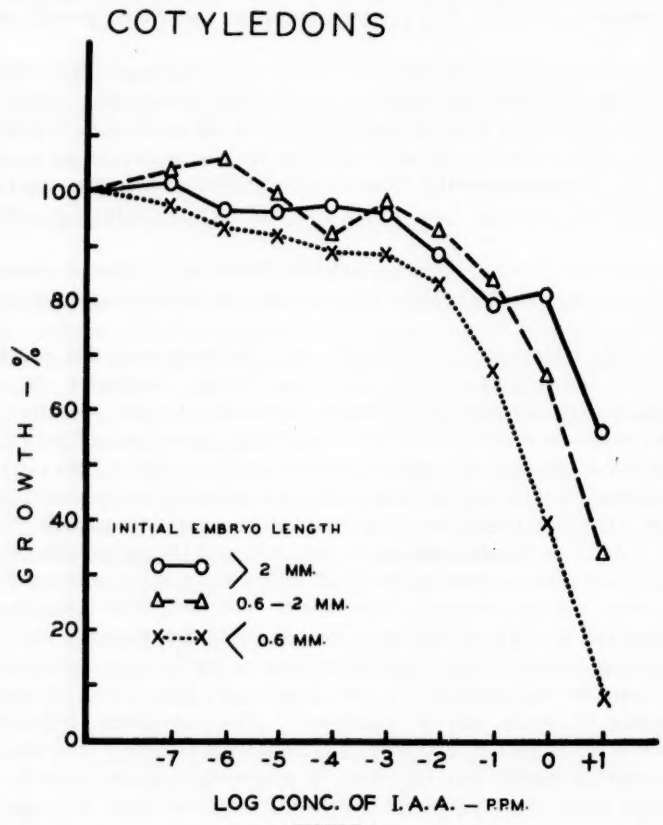


FIGURE 1

Relation between IAA concentration and growth of cotyledons of *Datura stramonium* embryos.

shows, the differences between the hypocotyl growth of differently sized embryos at 10 p.p.m. is not as great as between those of the cotyledons at this concentration. The growth values are lower and the hypocotyls of the smallest embryos do not grow, though the cotyledons showed some elongation (cf. figure 1: 10 p.p.m.). The hypocotyls of the larger embryos do

grow, but they look abnormal, are swollen and brittle. The cotyledons of these embryos, however, look normal.

At the lower concentrations some inhibition can be observed at 10^{-2} p.p.m. in the case of nearly mature and very small embryos, but lower concentrations do not seem to have any effect. A slightly beneficial

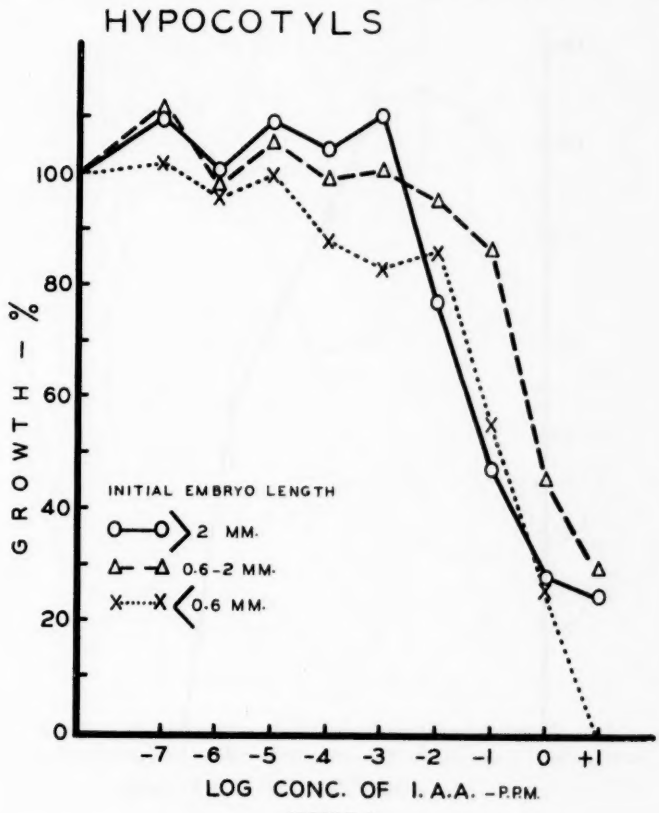


FIGURE 2

Relation between IAA concentration and growth of hypocotyls of *Datura stramonium* embryos.

effect seems to be caused by IAA in concentrations as low as 10^{-3} down to 10^{-7} p.p.m , but these effects are very small and hardly beyond the limit set by the experimental error. The minimum at 10^{-6} p.p.m. could not be explained.

In contrast to cotyledons and hypocotyls, the roots are not developed

at the beginning of the experiments, although the radicle is histologically well differentiated. In no case did any visible root system develop on the embryo inside the seed. Root formation in vitro was observed only when the hypocotyl attained a length of at least 3–4 mm. Embryos smaller than

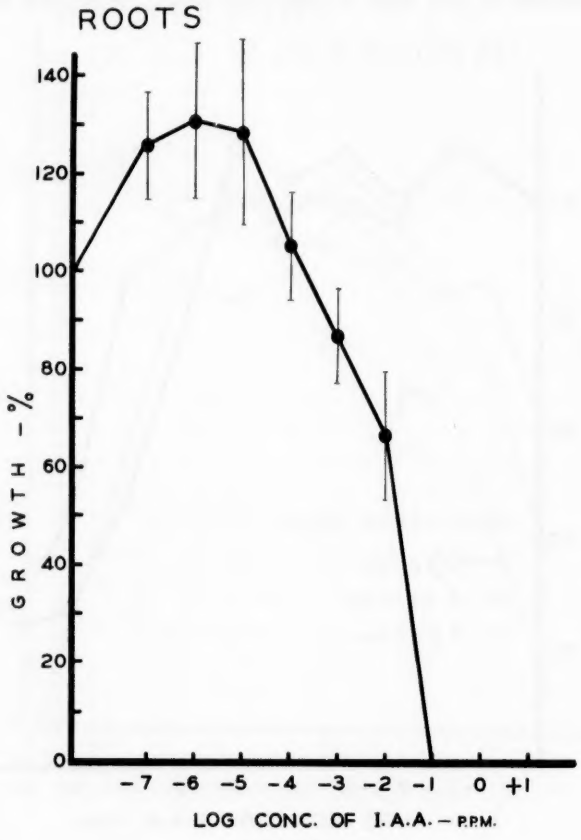


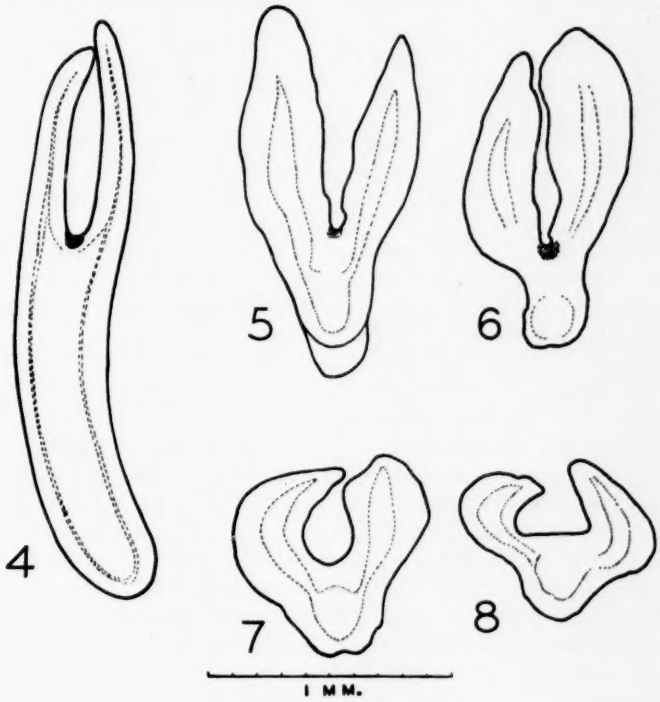
FIGURE 3

Relation between IAA concentration and growth of roots of *Datura stramonium* embryos.

0.6 mm. seldom formed roots during the experimental period and our observations on root growth therefore are limited to embryos longer than 0.6 mm. at the moment of excision from the ovules.

Figure 3 shows the relation between root growth and IAA concentration. Root formation is completely inhibited at 0.1 p.p.m.; 10^{-2} and 10^{-3} p.p.m.

were found to decrease the growth with an average of 35 and 15 per cent respectively. When the auxin concentration is still further decreased a stimulation becomes evident within a range from 10^{-5} to 10^{-7} p.p.m. This is in accord with the generally observed fact that root growth is stimulated by these extremely low concentrations.



FIGURES 4–8

Outlines made with camera lucida of sectioned embryos of *Datura stramonium* after 9 days growth on a medium with 1 p.p.m. IAA. Fig. 4: Control, Figs. 5–8: treated embryos.

Morphological Effects.—Together with the inhibition of the longitudinal growth there is a considerable effect on the morphological and histological condition of the embryos, which is different in the different parts. This can be seen from a comparison of the embryos represented in figure 4 with those in figures 5–8. Each embryo was dissected at the heart stage. The embryo in figure 4 was kept as control and grew on the basal medium, the embryos in figures 5–8 grew on a medium to which 1 p.p.m. IAA was added. There is no visible effect on the lower terminal portion of the

hypocotyl where a well-differentiated radicle and a root cap are formed (Figs. 5–8). This lack of effect is probably due to the fact that the terminal meristem of the root was already formed and differentiated when the embryos were placed in the medium. The auxin apparently only prevents further development of the primary root system.

The effect on the rest of the hypocotyl in each embryo is very strong. In some embryos it arrested almost completely the growth in length (Figs. 7 and 8). Comparison with the control embryo (Fig. 4) shows that the total length of the latter is about 4–6 times larger than embryos grown on a nutrient with 1 p.p.m. IAA.

The influence of auxin is also seen in the complete lack of meristematic tissue which gives rise to the epicotyl in later stages (Figs. 7 and 8). In less affected embryos (Figs. 5 and 6) traces of rudimentary meristem were seen at the base of the cotyledons.

The effect on the cotyledons is expressed in another way. As we have seen before, they may grow as long as those of the controls (Figs. 5 and 6) but their outlines are irregular. In other embryos, cotyledons became curved due to the more intensive growth of cells on their abaxial side (compare figure 4 with figures 7 and 8). The width of all the cotyledons is more than twice as large as that of cotyledons of embryos grown on nutrient without auxin. This increase in width of the cotyledons is due apparently to the fact that the cells became larger in width than in length. They differ in this respect from the cells in the cotyledons of the control embryos. For instance, in the control embryos the cotyledons consisted of about 6–7 cell layers, the cells being 33–42 μ long by 16 μ wide. In the treated embryos there are 7–8 cell layers, the cells being 16–28 μ long by 28–33 μ wide (all measurements referring to fixed material).

The large embryos showed a less pronounced response. Though the hypocotyls were rather swollen and brittle, the cotyledons seemed unaltered, a condition which, as in the case of the radicles, may be due to the fact that they were at a later stage of differentiation at the time of dissection.

Experiments with Malt Extract.—For the experiments just described a nutrient was used to which casein hydrolysate had been added to provide organic nitrogen. The early work on Datura embryo growth by Van Overbeek, *et al.*,[18] and by Blakeslee and Satina[2] had been carried out with media to which coconut milk or malt extract was added. These media have the advantage of approximating more closely the natural environment of the embryos, coconut milk being a liquid endosperm and malt extract being derived from germinating barley kernels.

Therefore we performed a number of experiments with malt extract instead of casein hydrolysate. It was found that IAA inhibition was distinctly less in the presence of malt extract than in that of casein hydrolysate, as we learn from table 1.

If we assume that this may be due to a protective effect of the malt against the auxin then we may expect that an increased malt concentration would bring about a decreased effect of the auxin. That this is actually the case is clearly demonstrated by table 2, showing that when the malt concentration is increased from 0.1 to 3 per cent no significant increase in growth occurs, but that the growth inhibition by IAA lessens. Effects similar to that of malt have been found by Van Overbeek[17] and McLane.[9] McLane[9] found that corn endosperm caused no inhibition of immature embryos but IAA in the same concentration as it appears in the endosperm caused a very strong inhibition of growth. Earlier Van Overbeek[17] described similar experiments with coconut milk. Pure coconut milk caused a very slight

TABLE 1

GROWTH OF EMBRYOS ON MEDIA WITH 0.45 PER CENT MALT EXTRACT OR 400 P.P.M. CASEIN HYDROLYSATE AFTER ADDITION OF 0, 1, OR 10 P.P.M. INDOLE-3-ACETIC ACID

Growth is expressed as relative growth, i.e., final length divided by initial length. Initial length 0.4 mm. (heart stage).

IAA,	CASEIN HYDROLYSATE		MALT EXTRACT	
	RELATIVE GROWTH	%	RELATIVE GROWTH	%
0	7.5 ± 1.2	100	6.0 ± 0.3	100
1	3.3 ± 0.4	44	3.3 ± 0.2	55
10	1.3 ± 0.1	17	2.5 ± 0.2	41

TABLE 2

GROWTH OF EMBRYOS ON MEDIA WITH DIFFERENT AMOUNTS OF MALT EXTRACT AND WITH OR WITHOUT 1 P.P.M. INDOLE-3-ACETIC ACID

Growth is expressed as relative growth, i.e., final length divided by initial length. Initial length 0.4 mm. (heart stage).

MALT CONCENTRATION, %	RELATIVE GROWTH	
	CONTROL	IAA
0.1	6.1 ± 0.4	2.4 ± 0.2
0.5	6.5 ± 1.0	3.2 ± 0.1
1.0	6.1 ± 0.7	3.5 ± 0.1
3.0	5.2 ± 0.5	3.8 ± 0.2

curvature with the Avena test, but an ether extract of the milk had a very high auxin content. He suggested that auxin may be combined with a substrate and thus be inactivated.

Some preliminary determinations of the auxin in endosperm of *Datura stramonium* ovules, by placing the endosperm on agar blocks and testing the auxin content of the agar, gave values as high as 0.13 μg. per gram of endosperm.

Discussion.—Our experiments have shown that IAA has a very strong inhibitory action on embryos of *Datura stramonium* in vitro. At the same time it may act as a growth stimulator on roots in very low concentrations. Growth stimulation of hypocotyls has been observed at very low concen-

trations, but it was non-significant. This different effect of different concentrations may explain some of the variable results of Tukey[16] and the contradictory statements made by Spoerl[14] and Curtis and Nichol,[3] the former claiming growth promotion, the latters growth inhibition.

The fact that IAA inhibits growth in vitro gives rise to the problem whether auxin in the endosperm affects the growth of the embryos. As has been mentioned before, endosperm of *Datura stramonium* ovules contains auxin Luckwill[8] and Hatcher[6] followed the auxin content of endosperm of apple and barley throughout the ontogeny of the seeds. In both endosperms the auxin content increased gradually to a peak value after which a rapid decrease occurred until maturity. In the apple seeds a previous optimum was found coinciding with endosperm formation.

Luckwill[8] suggested that the hormone plays a part in the embryo development and is used up more quickly than it is produced, thus causing a low auxin content before the second optimum. Hatcher[6] denies that there is a relation between auxin content of the endosperm and embryo growth. He points to the fact, however, that the growth rate of the embryo decreases when the auxin content is still increasing. It may yet be possible that the decrease of the embryo growth can be partly ascribed to the high auxin content of the endosperm.

The peculiar morphological condition brought about by indole-3-acetic acid on embryos observed in vitro is never found inside Datura seeds of selfs. It is likely that some factor in endosperm and also in malt regulates the auxin inhibition. This would mean that the embryo growth inside the ovule depends on a balance between inhibiting and promoting factors. Whether the effect of these controlling factors is direct or indirect on the auxin is an open question. At any rate, when this neutralizing agent is not present in in vitro cultures a growth inhibition follows caused by the auxin.

Summary.—The growth of *Datura stramonium* embryos in vitro is inhibited by indole-3-acetic acid in concentrations of 10^{-2} p.p.m. or higher. Young embryos are more sensitive than advanced embryos. The effect of auxin is differential: hypocotyl growth is more reduced than cotyledon growth. The morphological and histological responses of cotyledons and hypocotyls are also different. Roots are inhibited by 10^{-3} p.p.m. and higher concentrations. Root stimulation occurs from 10^{-7} to 10^{-5} p.p.m. Endosperm is found to contain auxin, which in vivo does not visibly affect embryo growth. Malt extract partly neutralizes the inhibitory effect of auxin in vitro.

* This work has been supported in part by a grant from the National Cancer Institute of the U. S. Public Health Service. Contribution from the Department of Botany, Smith College, New Series No. 47.

[1] Avery, G. S., Creighton, H. B., and Shalucha, B., *Am. J. Bot.*, **27**, 289 (1940).

[2] Blakeslee, A. F., and Satina, S., *Science*, **99**, 331 (1944).
[3] Curtis, J. T., and Nichol, Marion, A., *Bull. Torrey Bot. Club*, **75**, 358 (1948).
[4] Gustafson, F. G., *Am. J. Bot.*, **26**, 189 (1939).
[5] Guttenberg, H. von, and Lehle-Joerges, E., *Planta*, **35**, 281 (1947).
[6] Hatcher, E. S. J., *Ann. Bot.*, N.S. **9**, 235 (1945).
[7] Loo, S. W., and Wang, F. H., *Science*, **98**, 544 (1943).
[8] Luckwill, L. C., *J. Hort. Science*, **24**, 32 (1948).
[9] McLane, S., cited by A. E. Murneek in *Plant Growth Substances* (F. Skoog, ed.), Univ. Wisconsin Press, p. 333 (1951).
[10] Pohl, R., *Planta*, **25**, 720 (1936).
[11] Rietsema, J., Satina, S., and Blakeslee, A. F., *Am. J. Bot.;* (in press).
[12] Rijven, A. H. G. C., *Acta Bot. Neerlandica*, **1**, 157 (1952).
[13] Siu, R. G. H., personal communication.
[14] Spoerl, E., *Am. J. Bot.*, **35**, 88 (1948).
[15] Sterling, C., *Bot. Gaz.*, **111**, 90 (1949).
[16] Tukey, H. B., *Ibid.*, **99**, 630 (1938).
[17] Van Overbeek, J., *Cold Spring Harbor Symp. Quant. Biol.*, **10**, 126 (1942).
[18] Van Overbeek, J., Conklin, Marie E., and Blakeslee, A. F., *Am. J. Bot.*, **29**, 472 (1942).
[19] Voss, H., *Planta*, **30**, 252 (1939).

ON THE ORIGIN OF CONTINENTS AND MOUNTAINS

BY HAROLD C. UREY

DEPARTMENT OF CHEMISTRY AND INSTITUTE FOR NUCLEAR STUDIES, UNIVERSITY OF CHICAGO

Communicated June 26, 1953

In attempting to marshal evidence relative to the origin of the continents and mountains of the earth the attention of students of this subject has mostly been limited to the earth itself. Though most of the evidence must originate in data from the earth, additional information is supplied by the moon and Mars. If there were only several planets near enough to the earth for detailed observation, it seems likely that we could get a much more certain appraisal of the causes of mountain formation and the physical reasons for the existence of continents.

During the past few years, I have attempted to discover objective evidence relative to the physical conditions under which the earth originated, as of course have many others. Several lines of evidence which at first sight seem unrelated to each other can be used to construct an outline of events during the formation of the earth and these indicate the source of energy and the mechanism which produced the continents and mountains. In this paper a brief review of some observations and theoretical discussions by different men is given in approximately the order of publication and

following this there is presented an attempted synthesis of the evidence into one sequence of events.

1. The Convection Theory of Folded Mountains.—The contraction theory for the origin of folded mountains assumes that a rigid surface of the earth formed during its very early history and that the interior of the earth has been losing heat by conduction since then, and that the resulting contraction has caused a crumpled surface.[1] However, there appears to be no evidence to support the view that the earth's mean temperature is decreasing. The very open structure of fissures on the moon is most easily explained as due to increasing temperature in its interior and thus due to expansion. The interior of the moon should lose heat by conduction much more readily than the earth because of its smaller size. A reorganization of the constituents of the earth, i.e., the formation of the core, should produce higher temperatures within the earth, and a continuous growth of the core would produce increasing pressures and thus a decreasing radius with time. On the other hand, such a formation of a core would convert gravitational energy into heat and hence raise the temperature and cause an expansion of the earth. Thus the problem becomes much more complex than that assumed for a simple cooling of a spherical earth. Continuous erosion and thus degradation of the land masses requires a continual elevation of these masses by some effective mechanism and supply of energy, and if the water of the earth's surface has increased with time[2] an even greater demand is put on such a mechanism and energy supply.

Holmes[3] has proposed that convection currents within the earth are responsible for the formation of mountains. He suggests that rising currents exist in the Arctic region, under South Africa and the neighboring oceans, and under the Pacific Ocean, and that sinking occurs at the boundaries of these convection cells at the great mountain areas of the earth. Meinesz,[4] while he agreed with Holmes' ideas in general, maintained that the rising currents are under the continents and that sinking occurs under the oceans. He argued that high density areas exist under the oceans because of the generally positive gravity anomalies in these regions and that these high density areas must be sinking. Griggs[5] has made model experiments which show the successive formation of geosynclines, island arcs, and mountains in a very realistic way. His experiments agree with Holmes' postulates in regard to the regions of rising and sinking currents. It seems that Holmes' arrangement of currents must be assumed, if the convection hypothesis is to account for the formation of mountains and the growth of continents. The continents are obviously degraded by erosion, and convection currents can elevate them by moving material toward them, while currents moving away from the continents should remove material from under them just as pictured by Holmes. The convections postulated

should constitute an adequate mechanism for continent growth and maintenance and for mountain building.*

An important question is whether convection across the level at about 900 km. depth is possible. The density below the Mohorovičić discontinuity below the crust is usually assumed to be about 3.32 g. cm.$^{-3}$ as given by Bullen.[6] However, if the temperature is about 1000°C. and the coefficient of expansion is 4×10^{-5}, the material has a density of 3.45 at ordinary temperatures. This is the density of cosmic material with its iron completely oxidized,[7] or it is the density of a rocky material similar to cosmic material but containing less iron oxide together with some metallic iron, and is approximately the density of the moon corrected for temperature in the same way, which Urey[8] argues is composed of cosmic material with its iron oxidized. Between some hundreds of kilometers and about 900 km. depth, the density rises due to change of crystalline phase[8a] or to admixture of metallic iron or both. If the chemical composition is the same throughout the mantle, then convection should be possible throughout the mantle in spite of phase changes, provided that these changes occur rapidly as compared to the rate of convection, and the convection might well occur in either layer separately. But if additional metallic iron is present below some level then convection would occur only within the separate regions of uniform chemical composition. No evidence relative to this question seems conclusive at this time, but in some way the difficulties in understanding the complexities of mountain formation may be bound up with this question.

2. Mars and the Moon Do Not Have Cores or Folded Mountains.— The Clairaut constant is defined by the equation

$$\frac{\epsilon GM}{a^3\Omega^2} = \text{Clairaut constant}$$

where ϵ is the oblateness, i.e., the difference in equatorial and polar radii of the plant divided by the radius, G the gravitational constant, M, the mass, a, the radius, and Ω, the angular velocity of rotation of the planet. The value of this constant depends on the distribution of density within the planet, and is 1.25 for uniform density and 0.50 if all the mass were concentrated at the center of the planet. For other distributions the value is intermediate and its value for the earth is 0.98.

The application of this formula to Mars has been discussed by Urey.[9] The value of ϵ is determined from the motion of the moons which agree on the value, namely 0.005205 ± 0.000015, and hence this quantity is well known. The value of G is 6.670×10^{-8} The angular velocity is known with high precision. The value of the radius is least well known Urey selected Trumpler's value for this, namely 0.520 of the earth's radius, for reasons given by him. Since then Kuiper[10] has published a new radius of

0.523 of the earth's radius determined by visual comparison of the disk of Mars with an artificial disk of the same color and intensity and of adjustable diameter projected in the focal plane of the telescope near the image of the planet. These determinations of Kuiper appear to be the most reliable of any in the literature. However, Mars does have some atmosphere and hence Kuiper's value should be too large by some amount. The value of Trumpler still seems reasonable for the radius of the solid surface though the difference between these values is surely within Trumpler's error. In fact the radius of the solid surface might be less than that given by Trumpler. The value of the constant calculated from the observed data is 1.22, that is, near enough to 1.25 to suggest that Mars is a planet of uniform chemical composition throughout. Kuiper's value for a gives 1.19 and a slightly smaller value than Trumpler's would give a value of 1.25. Assuming that the density follows the formula $\rho = \rho_0 + \rho_1 p$, and that ρ_1 can be estimated from Bullen's values for the variation of density with pressure within the earth, Urey secured for the density,

$$\rho(r) = 4.02 + 2.05 \times 10^{-12} [{}^2/_3\pi G\rho_m^2 (a^2 - r^2)]$$

The quantity in brackets is the pressure as a function of r, where G is the gravitational constant, ρ_m, the mean density, 4.24 g. cm.$^{-3}$, and a, the radius of the planet, 3313 km. His value for the Clairaut constant calculated from theory for a planet of uniform composition with density following the above equation is 1.22. It appears that Mars must have no core, or a very small one at most.

Mars appears to have no mountains higher than 750 meters,[11] though elevated plateaus are indicated by some irregularity of the ice caps. It appears that mountains formation as observed on earth does not occur on Mars. Also, the dark markings on the planet are not particularly limited to one hemisphere as is found to be true for terrestrial continents, but show an equatorial distribution mostly a little south of the equator.

The moon's surface is remarkably uncompressed and indeed has many large and open cracks in its surface, and there are no evidences for folded mountains. The mean density is 3.34 g. cm.$^{-3}$ but the interior is probably at a higher temperature than the surface and hence its density at earth surface temperatures may be considerably higher. Urey's estimated value of 3.47 g. cm.$^{-3}$ is based on an assumed temperature of 1200°C. and a coefficient of expansion of 4×10^{-5}. This may not be correct and in fact is probably too high. However, some metallic iron may exist within the moon. The difference in moments of inertia about the moon's axis of rotation and about an axis in the direction of the earth is greater than the equilibrium value as indicated by its mechanical motion,[12] and hence the moon is not in isostatic equilibrium and probably has formed no core even

if metallic iron is present in its substance. Again there is a correlation between the absence of a core and the absence of folded mountains.

3. The Energy of Formation of the Earth's Core.—There has been very considerable discussion as to whether the earth's core was formed as the earth formed or whether it has been produced during geologic time. The proponents of the first view have generally postulated a high temperature molten primitive earth, while the latter view has been held in various forms by those who favor a low temperature for the earth's original state. If the core formed during geological time, a substantial amount of gravitational energy was converted to heat, a point which has been neglected.

Urey[13] has estimated the energy of formation of the earth with its present core from a primitive earth of uniform composition without a core. This estimate was approximate because of the many unknown quantities relating to such a primitive earth, the temperature, the distribution of density, the chemical composition, that is, the proportion of metallic, troilite and silicate phases, the possible crystalline modifications, etc. The estimated energy of formation of the present earth is 2.5×10^{39} ergs, and the energy of core formation 1.7×10^{38} ergs. The error of this last value may be taken as about ± 30 per cent. The rise in temperature due to the conversion of this energy to heat if none is lost from the earth is greater than 2000°C., which is more than the rise in temperature of the earth due to radioactive heating in 3×10^9 years.

If the earth formed at low temperatures separation of the core as it formed was not possible. It cannot be expected that the earth formed with completely uniform chemical composition throughout. In particular, the proportions of the metallic and silicate phases probably varied with the radius and to some extent with latitude and longitude. As pointed out previously,[14] small differences in density should have caused sinking of more dense parts of continental dimensions at fairly rapid rates. During such movements heat would be produced, temperatures would rise, viscosity decrease, and the movement would accelerate. The stage could have been set for a catastrophic reorganization of the earth. If the metallic and silicate phases separated by any means—and the solution mechanism previously suggested[15] seems most probable—unstable distributions of the phases could be produced continuously and thus maintain convection. It is only necessary that the rate of separation of the phases be greater at greater depths in order that lower-density material be generated at lower levels of the mantle, and thus an instability be maintained. If the separation of the phases is slow as compared to the convection cycle, a quasi-periodic convection might well result, thus accounting for a quasi-periodic production of mountains.

4. Convection in the Earth.—Meinesz[16] has given a very significant discussion of convection currents in the earth on the basis of Prey's[17] develop-

ment of the earth's topography in spherical harmonics. Meinesz took the square root of the sum of squares of the amplitudes for the spherical harmonics of each order for the deviation in thickness of the sialic crust of the earth, which he derived from Prey's analysis, and plotted these against the orders of the harmonics. This curve is reproduced in figure 1. There are distinct maxima and minima in the curve but the prominence of the first five harmonics is especially interesting. Prey's analysis represents the smoothed outlines of the earth's surface very well, and these regularities probably represent over-all processes for the formation of the earth's surface.

Meinesz ascribes the first-order harmonic, which is due to the land areas being on one hemisphere and the Pacific Ocean being on the other, to convection of the entire earth in a single cell during its early history. In the single-cell convection the motion is along one diameter through the earth and back along the surface. This convection is assumed to have moved the continents from the whole surface of the earth to a single continent on one hemisphere above the area of sinking of the convection current. He ascribes the prominent second, third, fourth, and fifth harmonics to later convections in smaller cells. He assumes that the core of the earth formed during the first single-cell convection and he shows, as does Jeffreys,[18] by approximate calculations that the single-cell convection would have been impossible if the earth's core had formed. Chandrasekhar[19] has given exact solutions for this problem and shows that convection in a single cell was impossible if a liquid core of greater radius than 18 per cent of the radius existed. Also he shows that the convections in two, three, etc., cells should have appeared as the core grew. His calculations, as are those of Jeffreys and Meinesz, are based upon the assumption that thermal energy, that is, the generation of heat in the interior due presumably to radioactive substances, produced the convections. As was pointed out by Kelvin, the number of cells in a convecting medium is determined fundamentally by the condition that breadths and heights of such cells are approximately equal, and hence the conclusions of these authors will apply very closely regardless of the

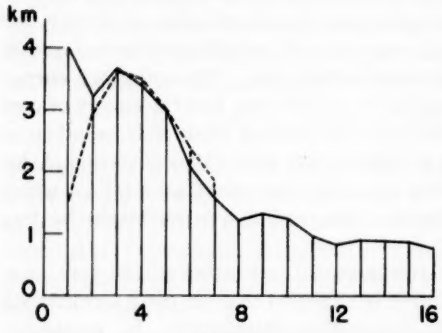


FIGURE 1

Meinesz' plot of the thickness of the crust as a function of the order of spherical harmonic amplitudes.

source of energy which drives the convections. Specifically, if the source of energy was chemical differentiation, that is, the formation of the core, it can be expected that convection in a single cell would not be possible if the radius of the earth's core was larger than about 18 per cent of its radius. Meinesz assumes that the later convections partially disrupted the single continent formed and pushed the fragments to about their present positions without scattering them uniformly over the earth's surface. He assumes that the core was produced by and during the single-cell convection, and that subsequent convection cells occurred only in the mantle and were produced by the cooling of the mantle. These he postulates continued intermittently with periods of the order of some hundreds of millions of years and have produced the folded mountains.

This discussion fails to consider the large energy liberated by the formation of the core, discussed above. This is potential energy and hence is more available than heat energy for mechanical work since it is not subject to the limitations of the second law of thermodynamics. If the metallic iron-nickel flowed to the earth's center to form the core, the temperature should have increased and the thermal convection theories do not apply and, in fact, the theory becomes very complex indeed, probably so complex as to defy analysis. Using 1.7×10^{38} ergs, the temperature rise becomes,

$$\frac{\Delta E}{cM} = \frac{1.7 \times 10^{38}}{1.25 \times 10^{7} \times 5.976 \times 10^{27}} \cong 2300°,$$

where c is the heat capacity per gram and is taken as 1.25×10^{7} ergs g.$^{-1}$ (0.3 cal. g.$^{-1}$). If a smaller value for the energy is taken, say 1×10^{38} ergs, still a large rise in temperature must result.

Urey[20] pointed out the importance of the gravitational energy, and that the earth must have formed with a high viscosity, and hence a low temperature sufficient to prevent the sinking of metallic iron-nickel and sulfide to form the core during the time of formation of the earth, if the single-cell convection caused the differentiation into land and water hemispheres. The conditions are that the earth had no core, or only a small one, and that primitive continental material had differentiated on the surface, and hence the earth was completely formed at the time of the first convection in a single cell. He also pointed out that rapid formation of the core would have raised temperatures, and if the energy were lost by conduction to the surface, in one-tenth of geologic time high thermal gradients would be required, and that a melted condition would have existed close to the earth's surface and would thus probably destroy the continents. He concludes that only a liberation of the core energy over a considerable time, and a considerable storage of heat, i.e., rise in temperature, can be expected to meet Meinesz' conditions. This argues for a low temperature during the accumulation of

the earth, followed by a generation of heat as the core formed and a rise of temperature in some time to its present value. The alternative to this conclusion is that the Meinesz suggestion relative to the origin of the water and land hemispheres must be discarded and this distribution regarded as "accidental." For many years it has generally and confidently been regarded otherwise, namely that this distribution was due to the loss of the moon from the region of the Pacific Ocean, although this theory has now been generally abandoned.

5. Rudimentary Continents on the Moon.—The terrestrial continents contain no metallic iron or nickel, little oxidized nickel and little iron sulfide. Their substance has been evidently completely molten at one time, either as a whole early in the earth's history or bit by bit during the lava flows since then. Original continental masses should have been extensive pools of magma through which metallic iron and nickel and iron sulfide sank. If the earth was formed in a completely molten state, they sank to form the core. If not, then they sank to a lesser depth and drained by some mechanism to the core.

The moon has had such pools of lava on its surface and they appear to be primitive. The collision maria, Maria Imbrium, Serenitatis, Crisium, Nectaris, Humorum, and perhaps Fecunditatis belong in this group. Gilbert[21] and others since him have concluded that these lavas were formed at the terminal stage of the moon's formation. It has been estimated[22] that they were 9 to 15 km. deep and that they were formed within a brief period of time, $\sim 10^4$ to 10^5 years. At the bottom of these lava pools, there should have been molten iron-nickel and iron sulfide.

The moon appears not to be in isostatic equilibrium, as pointed out above, and if isostatic equilibrium has not been attained it seems probable that iron-nickel and iron sulfide have not moved to the center of the moon. On the basis of Holmes' and Meinesz' postulates this would have occurred had metallic iron and iron sulfide been present at the bottom of silicate masses on the earth similar to the lunar maria, and the remaining silicate masses would have been quite adequate primitive continents floating on the primitive mixture of undifferentiated cosmic material.[23]

Gilbert argued that the collision maria resulted from collisions of satellites of the earth-moon system with the moon's surface. The escape energy of the moon is about 2.8×10^{10} ergs $g.^{-1}$ and hence is sufficient to melt the colliding object and some of the moon's surface materials, and so molten pools could be expected to occur on the moon. The escape energy of the earth is 22 times as great as that of the moon and the energy of the circum-surface orbit is one-half of this. Hence vaporization of objects falling on the earth is to be expected. The energy of the smaller ones would be dissipated in the atmosphere, but the collisions of the larger ones should have produced silicate rain storms and molten pools of magma. If these colli-

sions occurred sufficiently frequently the earth's surface would be completely molten, but if they occurred infrequently general temperatures would be low. Thus the processes which produced the lava maria of the moon also may have produced similar bodies on the earth.

Such primitive continents would have the mean density of cosmic matter except for the absence of metallic iron-nickel alloy and iron sulfide and hence a density of about 3.3 g. cm.$^{-3}$, but with less dense silicates at the surface and more dense ones at the bottom.[24] They would float in the solid mixture of "cosmic material" below. But if the iron-nickel and iron sulfide of the mantle moved downward to form the core, the material of the primitive continent would have the same density as the material surrounding it. Hence differentiation of silicates within the primitive continent, below it, and around it must have occurred. By partial fusion and other methods the sialic materials have become differentiated and primitive lunar type maria can only be assumed to have been nuclei about which the subsequent differentiation took place.

6. The Temperature of Formation of the Earth.—Bowen and others[25] at the Rancho Santa Fe conference pointed out that the abundance of silica at the earth's surface is lower than could be expected if the earth had solidified from a molten state. Briefly the argument is that solidification would begin at the core since freezing points of solids increase more rapidly with pressure than does the adiabatic temperature.[26] Olivines crystallize first, leaving sialic material to crystallize later. Hence olivines would crystallize deep in the earth leaving most of the silica and alumina for the surface regions. It may be that the fractionation observed at low pressures does not obtain at high pressures such as exist deep in the earth, and hence the argument becomes uncertain.

Toward the end of the solidification process of a molten earth the cooling process should have produced crystals sinking in the liquid and these should have exchanged with the liquid to some extent as they sank. This would have produced a counter-current fractionation process similar in principle to distillation columns. If the crystals were not too large for efficient exchange, many "theoretical plates" should have been available in the outer 100 km. of the earth's surface and hence very high concentrations of sial should have been produced at the surface. (Parenthetically, tektites have a composition such as might have been produced in this way and Nininger's[27] suggestion that they originated from the moon's surface where such processes evidently occurred has a certain plausibility.) The absence of large quantities of such sial on the surface parts of the earth argues against such a completely molten primitive earth.

Recently, Urey[28] has shown that the elements which are concentrated in oceans and sediments as compared to igneous rocks are either volatile at ordinary temperatures, i.e., H_2O, carbon as CH_4 or CO_2, nitrogen as N_2

or NH_3, or soluble in water at these temperatures, i.e., chlorine as chlorides bromine as bromides, boron as boric acid or borates, carbon slightly as carbonates and bicarbonates, and nitrogen as ammonia and ammonium salts, and that elements which should be volatile at a few hundred degrees centigrade such as mercury, arsenic, and cadmium in a reducing atmosphere are not concentrated at the earth's surface. From known thermodynamic properties of compounds of these elements he concludes that the prevailing temperatures during the latter part of the earth's formation were 100–200°C. He also shows that it is difficult to understand the occurrence of nitrogen in igneous rocks unless it was buried as ammonium salts or as organic compounds containing nitrogen, and this again indicates formation at low temperatures. The argument shows that the existence of extremely high temperatures generally during the terminal stages of formation is inconsistent with the distribution of volatile elements in the earth's surface.

This conclusion does not exclude the possibility of high temperatures over limited areas for limited times. Magma pools such as those observed to have existed on the moon are not excluded provided that they did not cover the earth completely, since volatile elements vaporizing in a high-temperature region could condense in the low-temperature regions and thus be buried in the accumulating material. However, if the single-cell convection discussed above is responsible for the continents being present in only one hemisphere, substantially the entire earth must have accumulated at low temperatures.

The mean temperature of the earth's interior was surely not as low as the surface temperature. Large objects falling on the earth would generate heat below the surface, and this could be lost only by conduction to the surface or by the circulation of liquid water, if liquid water was present on the earth at that time. Thus while a temperature of boiling water is reasonable for the surface, some hundreds of degrees is more reasonable for subsurface temperatures.

Such a low temperature at the time the earth accumulated means that the entire energy for the formation of the core could have been liberated in the early history of the earth without temperatures rising beyond those currently assumed for the interior of the earth.

7. *The Indicated Sequence of Events.*—The earth was formed by the accumulation of substantial objects which were satellites of the earth-moon system and which contained metallic iron-nickel alloy, iron sulfide and silicates as major phases. The iron and associated elements had previously been concentrated relative to the lithophile elements.[29] The rate of accumulation was such that low temperatures, ~100–200°C or less, prevailed at the surface and somewhat higher temperatures at greater depths. At this time the earth was a mixture of iron-nickel alloy, iron sulfide, and silicate minerals, but probably was not entirely uniform in composition with

respect to radius or latitude and longitude. The material was highly viscous, and sinking of more dense parts did not occur appreciably during the time of formation. Some differentiation occurred at the surface by melting and sinking of iron-nickel and iron sulfide in magma pools, and these areas formed primitive continental shields similar to those of the collision maria of the moon.

Convection occurred during a first epoch of the earth's existence in a single cell with more dense parts moving to the interior. The less dense surface areas floated to one hemisphere and formed one continent. This convection proceeded for about one cycle because the higher density material remained at the center of the earth. The large gravitational energy was converted to heat and raised the temperature throughout the earth to values approaching those of the present time. The iron-nickel and iron sulfide melted and began sinking through the silicates to form a consolidated core. Such separation produced high temperatures and chemical inhomogeneities which initiated new convection, which now, since a core had formed, took place in two cells and partially broke up the single continent to form the present continental masses. Following this, other convections added their effects to the spherical harmonics of the earth's surface. The principal energy driving these convections was the energy of core formation, though radioactive heating and loss of energy at the earth's surface contributed to the effects. These convections are continuing at the present time and constitute the principal machines for mountain building and continent maintenance. It is probable that some iron-nickel and iron sulfide remain in the mantle and that their slow movement to the core constitutes an important source of energy for mountain building.

Throughout this process, partial melting and diffusional granitization have caused differentiation of the continental regions and very great segregation of certain elements in the surface regions, e.g., silicon, aluminum, sodium, potassium and calcium and minor constituents, notably uranium and thorium. Some volatile substances and others soluble in water are notably concentrated in the oceans and sedimentary rocks These increased concentrations can be explained in the framework of the above outline of events, but no unique explanation of them follows from it.

Mars probably has not produced convection currents and folded mountains. This difference from the earth can be explained as due to its lower temperature of formation resulting from its smaller accumulation energy per gram and hence its higher viscosity, and also to its smaller size, smaller gravitational field and smaller proportion of iron, all of which act to decrease convection. It should be noted that convection in the earth did not occur during the period of formation but that it did in some fraction of geological time, and hence it follows that convection in Mars should not have occurred probably even up to the present time.

The theory of convection induced in a sphere by variations in density has not been solved. The material flows through a tube along a diameter of radius about $1/\sqrt{2}$ that of the sphere and back along an annulus with one free surface of about the same length. Gravity acts on the high-density material moving downward along the diameter and the low-density material moving upward, but not greatly on the horizontal outer flow. Approximately the velocity of flow should be,

$$v = \frac{1}{32}\frac{a^2}{2}\frac{g\Delta\rho}{\rho}\frac{1}{\nu},$$

where a is the radius of the sphere, and ν the kinematic viscosity. We assume that the time during which this occurred was 3×10^8 years or 10^{16} sec., and the distance moved before the energy dissipated decreased the viscosity to a much smaller value was $a/2$. Then

$$\frac{a}{2} = vt = \frac{1}{64}\frac{a^2 g\Delta\rho}{\rho}\frac{1}{\nu}t.$$

Taking $\Delta\rho/\rho$ equal to 0.25, $g = 10^3$, $a = 6.37 \times 10^8$, we secure 5×10^{25} for ν which is 1.6×10^4 times the present estimated value. In view of the well-known rapid change of viscosity with temperature this is not unreasonable, if the temperature of the earth has risen by 1500 or 2000°C. If the viscosity of Mars were ten times that calculated for the earth the time for the convection in that planet would be 15×10^9 years and thus the convection within the earth and its absence in Mars would be understood. The assumption that density inhomogeneities should be the same in the two planets is purely an *ad hoc* one. Eventually radioactive heating may decrease the viscosity of that planet until convection is possible. Similar considerations apply even more definitely to the moon.

This series of events is similar to one presented at the Rancho Santa Fe conference by the writer,[30] but which was received with considerable skepticism and hence was never published. Briefly the argument was as follows. The moon of low density appears to consist of material of cosmic composition with nearly all its iron in the oxidized state. It is assumed to have accumulated at low temperatures in the presence of condensed water and ammonia and hence contains mostly oxidized iron and little metallic iron, as expected at these temperatures. Other bodies of this kind, mostly smaller, existed in the earth's protoplanet, one of which formed the nucleus for the accumulation of the earth. Following this stage a high-temperature process occurred which volatilized the silicates to a greater degree than the iron, and these volatilized silicates were lost when the gases were lost from the protoplanet, leaving substantial objects containing increased amounts of iron relative to the silicates. In the subsequent accumulation the earth, starting with the nucleus of lunar-type material, acquired most of these

iron-rich objects because of its central position in the protoplanet and greater mass, and thus became an object of high density. The moon acquired only a few of these objects because of its non-central position, and hence its density is less than that of the earth and its composition approximately that of cosmic non-volatile matter. This proposed series of events to this point has been published and is subscribed to by the writer at the present time. However, it was suggested at the Rancho Santa Fe conference that this accumulation process produced an unstable earth with a low-density core and high-density iron-rich mantle. It was assumed that a convection occurred early in the earth's history which transferred the low-density oxidized core to the position of the outer mantle and the high-density iron-rich mantle to the earth's interior and moved a low-density sialic surface layer, assumed to have been produced by a generally molten surface condition produced during the growth of the earth, to one side of the earth to produce the land hemisphere. This is very similar to the series of events deduced in this paper. I believe that the outline of events presented at the Rancho Santa Fe conference was more reasonable than it appeared to be at the time. However, the lines of evidence presented in this paper present better arguments for the validity of the conclusions presented at that meeting. In particular, the great importance of the studies of Meinesz should be emphasized in this connection.

The extreme form of these former suggestions of the writer in regard to the origin and composition of the mantle illustrates one way by which inhomogeneity in the primitive earth could have occurred. If the outer mantle is indeed different in composition from the inner mantle, convection within the core, the inner mantle, and the outer mantle separately should be possible, but not throughout the mantle as a whole. This requires modification of the theories of Meinesz, Jeffreys, and Chandrasekhar. But since these theoretical discussions have not considered the energy of formation of the core, they should be regarded as illustrative rather than judged as entirely true or false. That all the complexities of mountain formation as presented by Bucher,[31] for example, could be explained by a simple theory of any kind seems to be inherently doubtful. Meinesz has maintained that Prey's higher harmonics are due to convections in smaller cells, and this suggests convections in thinner layers such as the inner and outer mantles.

* As pointed out to me by Dr. Vening Meinesz on a recent visit, a convection current seems to be flowing under great areas of Asia in order to produce the folding processes of the East Indies and the Himalayas.

[1] An excellent review of theories of mountain formation has been given by A. Holmes, *J. Wash. Acad. Sci.*, **23**, 169 (1933). See also, B. Gutenberg *"Internal Constitution of the Earth,"* Dover Publications, Inc., New York, 1951, particularly pp. 191 ff.

[2] See Rubey, W. W., *Bull. Geol. Soc. Am.*, **62**, 1111 (1951) and references there cited.

[3] *Loc. cit.*

[4] Meinesz, F. A. Vening, *Koninkl. Akad. Wetenschap. Amsterdam*, **37**, 37 (1934).

[5] Griggs, D., *Am. J. Sci.*, **237,** 611 (1939).

[6] Bullen, K. E., *Mon. Not. Roy. Astron. Soc., Geophys. Suppl.*, **3,** 395 (1936); **6,** 50 (1950); *Trans. Roy. Soc. New Zealand*, **67,** 122 (1938); *Bull. Seis. Soc. Am.*, **30,** 235 (1940); **32,** 19 (1952). Numerical values are given in Urey, H. C., *The Planets*, Yale University Press, 1952, App. 4.

[7] See Urey, H. C., and Craig, H., *Geochim. et Cosmochim. Acta*, **4,** (in press) (1953). Their "low-group" abundances are used for reasons given by them.

[8] Urey, H. C., *The Planets*, Yale Univ. Press, 1952, p. 204.

[8a] See F. Birch, *J. Geophys. Res.*, **57,** 227 (1952).

[9] Urey, H. C., *The Planets*, pp. 62–66.

[10] Kuiper, G. P., Atmospheres of the Earth and Planets, 2nd ed., University of Chicago Press, 1952, p. 308.

[11] This is Lowell's estimate. E. C. Slipher recently expressed the view that Mars has no mountains higher than 2000 feet.

[12] Urey, H. C., *The Planets*, pp. 20–25.

[13] *Ibid.*, pp. 173–179.

[14] *Ibid.*, p. 181.

[15] *Ibid.*, pp. 90 and 91.

[16] Meinesz, F. A. Vening, *Proc. Koninkl. Ned. Akad. Vet.*, **53,** 7 (1950); **54,** 212–220 (1951); **55,** 527–553 (1952). *Jaarb. Kon. Ned. Akad. Vet.* (1951–1952). *Geol. en Mijnbouw* **14,** 11, 373–384 (Nov. 1952).

[17] Prey, A., *Abhandl. Ges. Wiss. Göttingen, Math.-physik., Kl. N. F.*, **11,** 1 (1922).

[18] Jeffreys, H., *Mon. Not. Roy. Astron. Soc., Geophys. Suppl.*, **5,** 272 (1952).

[19] Chandrasekhar, S., *Phil. Mag.*, **44,** 233 (1953).

[20] Urey, H. C., *Ibid.*, **44,** 227 (1953).

[21] Gilbert, G. K., *Bull. Phil. Soc. Wash.*, **12,** 241 (1893). See also Baldwin, R. B., *The Face of the Moon*, University of Chicago Press, 1949.

[22] *The Planets*, pp. 37 and 38.

[23] Urey, H. C., *Vistas in Astronomy*, (in press), Cambridge University Press, 1954, edited by A. Beer. In this paper a review of the evidence for the surface structure of the moon is given and the continental character of the collision maria is discussed.

[24] This is a calculated value for the "silicate phase," i.e., all constituents except FeS and metallic phase, for the "low-group" chondrites of Urey and Craig, *Geochim. Cosmochim. Acta*, **4,** (in press) (1953).

[25] See Slichter, L., PROC. NATL. ACAD. SCI., **36,** 511 (1950).

[26] Holmes, A., *loc. cit.*, shows diagrams indicating the possible relations. The work of Bridgeman, P., (*Proc. Am. Acad. Arts Sci.*, **77,** 187 (1949)) and Simon, F., (personal communication) shows that the relation is that given in the text.

[27] Nininger, H. H., *Sky and Telescope*, II, No. 16, 12; No. 17, 8 (1943).

[28] Urey, H. C., *Proc. Roy. Soc.*, (in press) (1953).

[29] See *The Planets*, Chap. 4, Section b.

[30] Urey, H. C., *Geochim. et Cosmochim. Acta*, **1,** 209–277 (1951). See, particularly, footnote, p. 248.

[31] Bucher, W. H., *The Deformation of the Earth's Crust*, Princeton University Press, 1941.

BOTTLENECK PROBLEMS AND DYNAMIC PROGRAMMING

BY RICHARD BELLMAN

RAND CORPORATION, SANTA MONICA, CALIFORNIA

Communicated by J. von Neumann, July 17, 1953

1. Introduction.—The purpose of this paper is to indicate how the theory of dynamic programming provides a mathematical formulation and a systematic approach to an interesting and significant class of production and allocation problems which we shall call "bottleneck" problems. The following two problems are typical:

Problem 1: At some initial period we possess a quantity x_1 of steel and a capacity x_2 of steel production, with the privilege of dividing the quantity x_1 into three parts, y_1, y_2, y_3, where y_1 is to be used to increase the capacity, y_2 to be used together with the present capacity to produce more steel, and y_3 to remain in the stockpile. Given that this allocation and production process continues for a fixed number of time periods and given the increase in capacity determined by x_1 and y_1 and the increase in the quantity of steel determined by x_2 and y_2, the question is to determine the allocation policy which maximizes the quantity of steel in the stockpile at the end of the final period.

Problem 2: We are engaged in the manufacture of an item which requires two component parts. We possess resources which may be used to increase the rates of production of these items. Assuming that we know the relation between rates of production and the rate of allocation of resources, what policy of allocation of resources do we employ in order to maximize the output of complete items over a fixed time interval?

These problems in their discrete forms where decisions are made at a finite number of fixed times may in some cases be attacked by the computational techniques of the theories of linear and non-linear programming. However, not only do the simplest problems, taken over time intervals of quite reasonable length, give rise to matrices of unreasonable dimensions but, more serious from the esthetic and mathematical point of view, this approach takes no cognizance of the intrinsic structure of the process.

If we consider the problem in its continuous form (an approximation technique which almost always results in a great analytic simplification), simple assumptions of linearity yield the system of differential equations

$$\frac{dx_1}{dt} = a_1 y_1, \qquad x_1(0) = c_1, \tag{1}$$

$$\frac{dx_2}{dt} = -y_1 - y_2 + \min(b_1 x_1, a_2 y_2) = (a_2 - 1)y_2 - y_1, \qquad x_2(0) = c_2,$$

with y_1 and y_2 subject to the restrictions

$$\begin{aligned} &(a) \quad y_1, y_2 \geq 0 \\ &(b) \quad y_1 + y_2 \leq x_2 \\ &(c) \quad y_2 \leq \frac{b_2 x_1}{a_1} \end{aligned} \tag{2}$$

The term $\min(b_1x_1, a_2y_2)$ represents the bottleneck condition and is the mathematical equivalent of the limitation on the growth of the steel stockpile due to insufficient capacity. Our aim is now to maximize $x_2(T)$.

In this form the problem may be attacked by classical methods of the calculus of variations utilizing Lagrange multipliers. However, once again this method is too general and the particular features of the problem are overshadowed.

A third attack on problems of this genre is afforded by the very interesting techniques developed by R. Isaacs in his study of a class of continuous games. Although there are some points of contact between our methods, the conceptual bases and the analytic continuations are quite distinct. In particular, his techniques would seem to lead to grave computational difficulties.

Let us mention that a simple interpretation of the second problem above leads to the mathematical question of maximizing $\int_0^T \min(x, y)\, dt$ over all f_1 and f_2 satisfying $0 \leq f_1, f_2 \leq M$, $\int_0^T (f_1 + f_2)\, dt \leq c_3$, where

$$\begin{aligned} \frac{dx}{dt} &= a_1x + a_2y + f_1, \qquad x(0) = c_1 \\ \frac{dy}{dt} &= a_3x + a_4y + f_2, \qquad y(0) = c_2. \end{aligned} \tag{3}$$

2. The Dynamic Programming Approach.—We shall consider only the first problem in discussing the application of dynamic programming techniques. (For a brief discussion of the theory and some typical problems, see refs. 1, 2.)

Our first step is to imbed the problem within a continuous family by taking the quantities x_1, x_2, and T, the duration of the process, as the basic state variables. The effect of any decision, where by a decision we mean a choice of y_1, y_2, and y_3, will be to transform these state variables into a similar triple. The criterion function max $x_2(T)$ may be written $f(x_1, x_2, T)$. Using the intuitive and obvious property that an optimal policy has the property that its continuation after any initial decisions must also be optimal with respect to the new state variables, we obtain the characteristic functional equation of dynamic programming

$$f(c_1, c_2, s + t) = \max_{D[0, s]} f(x_1(s), x_2(s), t),$$

$$f(c_1, c_2, 0) = c_2, c_1, c_2 \geq 0, \tag{1}$$

where x_1 and x_2 satisfy (1) of §1 and by $\max_{D[0, s]}$ we mean that the maximum is to be sought over all y_1 and y_2 which satisfy the restrictions in (2) of §2 over $[0, s]$. It is quite easy to prove by elementary arguments that the maximum is assumed. Furthermore, it follows readily from (1) that the solution is unique. For setting $t = 0$, we obtain, using the initial condition

$$f(c_1, c_2, s) = \max_{D[0, s]} f(x_1(s), x_2(s), 0) = \max_{D[0, s]} x_2(s). \tag{2}$$

Hence the functional equation is equivalent to the original process. Alternatively, it is easy to establish directly by successive approximations that (1) has a solution.

Assuming that f has well-behaved partial derivatives with respect to the basic variables, we obtain via a limiting process, as $s \to 0$, the partial differential equation

$$\frac{\partial f}{\partial t} = \max_{D[0]} \left[a_1 y_1 \frac{\partial f}{\partial c_1} + ((a_2 - 1)y_2 - y_1) \frac{\partial f}{\partial c_2} \right] \tag{3}$$

where by $D[0]$ we mean the region in y_1, y_2 space described by

$$0 \leq y_1, y_2, \qquad y_1 + y_2 \leq c_2, \qquad y_2 \leq b_1 c_1 / a_2. \tag{4}$$

Since it seems rather difficult to prove directly that the solution of (1) possesses the requisite continuity properties, an indirect approach is used. It is first shown without difficulty that any solution of (3) satisfies (2). Having obtained a solution of (1) by the use of (3), the uniqueness theorem tells us that it is the solution of (1).

Turning to the task of finding solutions of (3) we first consider the case of small t where the solution is immediate. Having obtained f for small t, the partial differential equation is now utilized to deduce the form of solution as t grows. Since this is a "bootstrap" method it is necessary to verify with great care that the f obtained in this fashion is actually a solution of (3). Having found f we can now determine whether or not the optimal policy is unique. It turns out to be so in this case.

3. An Outline of the Analytic Procedure.—Let us give a short sketch of the steps to be followed in obtaining the solution of the maximization problem. We shall consider the most complicated case where no capacity restriction exists initially, i.e., $c_2 < b_1 c_1 / a_2$. For processes of short duration the solution is clear, $y_1 = 0$, $y_2 = x_2$, yielding $f = c_2 e^{(a_2 - 1)t}$. This policy

is pursued until a bottleneck develops. Using the above allocation program, this will occur as soon as t exceeds $T_1 = (\log b_1c_1/a_2c_2)/(a_2 - 1)$.

In order to obtain the solution for times greater than T_1, we rewrite the right side of (3) of §3 in the form

$$\frac{\partial f}{\partial t} = \max_{D[0]} \left[y_1 \left(a_1 \frac{\partial f}{\partial c_1} - \frac{\partial f}{\partial c_2} \right) + (a_2 - 1) \frac{\partial f}{\partial c_2} y_2 \right]. \tag{1}$$

The location of the maximizing point $(y_1(0), y_2(0))$ will depend upon the sign and magnitude of the coefficients of y_1 and y_2. For $t < T_1$, the coefficient of y_1 is uniformly negative, while that of y_2 is uniformly positive.

Proceeding on the assumption of continuity, we suspect that the solution for t slightly greater than T_1 will have the form $y_1 = 0$, $y_2 = x_2$ for $0 \leq s \leq T_1$ and $y_1 = 0$, $y_2 = b_1x_1/a_2$ (due to the capacity restriction) for $T_1 \leq s \leq t$.

The function f will now have the form

$$f = \frac{b_1c_1}{a_2} + (t - T_1) \frac{(a_2 - 1)}{a_2} b_1c_1. \tag{2}$$

In order to find out how long this policy endures, we apply (3) of §2 at the point $s = T_1$. Starting from this point we see that f has the form $c_2' + (a_2 - 1)b_1c_1'(t - T_1)/a_2$.

The coefficient of y_1 in (1) is $a_1 \partial f/\partial c_1' - \partial f/\partial c_2' = a_1b_1(a_2 - 1)(t - T_1)/a_2 - 1$. This is 0 at $t - T_1 = a_2/a_1b_1(a_2 - 1) = T^*$. This shows that this modified policy is optimal for $T_1 \leq t \leq T_1 + a_2/a_1b_1(a_2 - 1)$.

For larger t, there will be an interior interval during which $y_1 = x_2 - b_1x_1/a_2$. This corresponds to allocating steel to increase the capacity of steel mills. For large t, the optimal policy has the form

$$y_1 = 0,\ y_2 = x_2,\ 0 \leq s \leq T_1$$

$$y_1 = x_2 - \frac{b_1x_1}{a_2},\ y_2 = \frac{b_1x_1}{a_2},\ T_1 \leq s \leq t - T^*$$

$$y_1 = 0,\ y_2 = \frac{b_1x_1}{a_2},\ t - T^* \leq s \leq t, \tag{3}$$

where T^* is as above.

Since we have pursued a bootstrap method it is essential as mentioned above that this solution be verified. With application to interindustry problems in mind, it is vital that this verification be analytic rather than numerical. The details of the verification are not trivial and we shall postpone any discussion until a future time.[3]

4. Discussion.—The method we have presented in skeleton form is applicable to a large class of problems related to the maximization of functions or functionals depending upon the solutions of differential

equations. It is also applicable to various classes of multistage continuous games, such as games of survival.

If—as is necessary in more realistic mathematical models dealing with the production of capital goods—time lags are taken into account, the complexity of the problem increases.

A complete and detailed treatment of the above problem will be presented subsequently, together with a discussion of extensions in the directions just cited.

[1] Bellman, R., "On the Theory of Dynamic Programming," PROC. NATL. ACAD. SCI., **38,** pp. 716–719 (1952).

[2] Bellman, R., Glicksberg, I., and Gross, O., "On Some Variational Problems Occurring in the Theory of Dynamic Programming," *Ibid.*, **39,** 298–301 (1953).

[3] Note added in proof: In the meantime, we have developed a method based upon simultaneous consideration of the problem and its dual, which makes the verification quite simple.

ON STEENROD'S REDUCED POWERS, THE INDEX OF INERTIA, AND THE TODD GENUS

BY FRIEDRICH HIRZEBRUCH

THE INSTITUTE FOR ADVANCED STUDY

Communicated by S. Lefschetz, July 3, 1953

Introduction.—This note is a preliminary report on some results concerning the Steenrod reduced powers[1] in *oriented* manifolds M^m and the index τ of such a manifold.

Definition of τ: If the dimension m of M is not divisible by 4, then $\tau(M) = 0$. If $m = 4k$, then $\tau(M)$ is equal to the number of positive squares minus the number of negative squares of the normalized quadratic form defined by the cup-product $x \smile x$ where $x \in H^{2k}(M^{4k}, R)$.

We apply the results to almost complex manifolds, in particular we give a definition of the Todd[2] genus of an almost complex manifold and state properties of this genus. Since a special algebraic formalism will be used throughout the note, we outline this formalism in the first section. All manifolds occurring in this note are compact, differentiable, and *oriented* unless stated to the contrary. Full details with further applications will appear elsewhere.

1. Multiplicative Γ*-Sequences.*—Let $\sum_{i=0}^{\infty} a_i x^i$, $a_0 = 1$, be the power series with the indeterminates a_i as coefficients, and let Γ be a field. Let $\{K_j\}$ be a sequence of polynomials, K_j being of weight j in the a_i and having

coefficients in Γ ($K_0 = 1$). If $C = \sum_{i=0}^{\infty} c_i x^i$, $c_0 = 1$, is an arbitrary power series, we denote by $K(C)$ the series $\sum_{j=0}^{\infty} K_j(c_1, c_2, \ldots, c_j)x^j$. We call $\{K_j\}$ a multiplicative Γ-sequence provided that K is a homomorphism, i.e., $K(AB) = K(A)K(B)$, where A and B are power series with indeterminates a_i, b_i as coefficients ($a_0 = b_0 = 1$). We construct for every given power series $Q(x) = \sum_{i=0}^{\infty} \gamma_i x^i$, ($\gamma_i \in \Gamma$, $\gamma_0 = 1$), a multiplicative Γ-sequence. Writing formally

$$1 + a_1x + \ldots + a_mx^m = (1 + \alpha_1x)(1 + \alpha_2x) \ldots (1 + \alpha_mx),$$

we express $Q(\alpha_1x)Q(\alpha_2x) \ldots Q(\alpha_mx)$ as a power series with coefficients which are polynomials in the a_i:

$$Q(\alpha_1x)Q(\alpha_2x) \ldots Q(\alpha_mx) = \sum_{j=0}^{\infty} K_{j,m}(a_1, \ldots, a_j)x^j.$$

One verifies easily that $K_{j,m}$ does not depend on m for $j \leq m$. We write $K_{j,j} = K_j$ and obtain the unique multiplicative Γ-sequence with $K(1 + x) = Q(x)$.

Now let Γ be the rationals. We denote by $\{T_j\}$ the multiplicative sequence belonging to $Q(x) = -x(e^{-x} - 1)^{-1}$ and call it the Todd sequence.[2]

$$2T_1 = a_1,\ 12T_2 = a_1^2 + a_2,\ 24T_3 = a_1a_2,\ 720T_4 = -a_4 + a_3a_1 + 3a_2^2 + 4a_2a_1^2 - a_1^4.$$

For a prime $q \geq 2$ the coefficients of $q^rT_{(q-1)r}$ are integers mod q (i.e., do not contain q in the denominators).

We also consider the multiplicative sequence belonging to $Q(x) = \sqrt{x}(\mathrm{tgh}\sqrt{x})^{-1}$ which we denote by $\{L_j\}$. We have

$$3L_1 = a_1,\ 45L_2 = 7a_2 - a_1^2,\ 945L_3 = 62a_3 - 13a_2a_1 + 2a_1^3, \ldots$$

For a prime $q \geq 3$ the coefficients of $q^rL_{1/2(q-1)r}$ are integers mod q.

2. *Reduced Powers.*[1]—Let M^m be a compact *oriented* manifold. The reduced powers are defined for every odd prime q:

$$\mathcal{P}_q^r: H^k(M^m, Z_q) \to H^{k+2(q-1)r}(M^m, Z_q).$$

In case $k + 2(q - 1)r = m$, there exists by Poincaré duality an element $s_q^r \in H^{2(q-1)r}(M^m, Z_q)$ such that

$$\mathcal{P}_q^r u = s_q^r u \quad \text{for all} \quad u \in H^k(M^m, Z_q).$$

Let $p_1, p_2, \ldots$ be the Pontrjagin classes[3] of M^m where $p_i \in H^{4i}(M^m, Z)$. In the notation of Wu[3] $p_i = P_0^{4i}$.

THEOREM 2.1. *The class s_q^r can be expressed as a polynomial in the Pontrjagin classes:*

$$s_q^r = q^r L_{1/2(q-1)r}(p_1, p_2, \ldots, p_{1/2(q-1)r}) \bmod q.$$

THEOREM 2.2. *If M^m is an almost complex manifold of n complex dimensions ($m = 2n$), then we can express s_q^r as a polynomial in the Chern classes c_i where $c_i \in H^{2i}(M^m, Z)$:*

$$s_q^r = q^r T_{(q-1)r}(c_1, c_2, \ldots, c_{(q-1)r}) \bmod q.$$

We have assumed that the prime q is odd, but in the case $q = 2$ we may consider the Steenrod squares Sq^i. For a manifold M^m (not necessarily oriented) Wu[4] defined the class $U^i \in H^i(M^m, Z_2)$ by $\mathrm{Sq}^i v = U^i v$ for all $v \in H^{m-i}(M^m, Z_2)$.

THEOREM 2.3. *Let $w_i \in H^i(M^m, Z_2)$ be the Stiefel-Whitney classes of M^m. We have*

$$U^i = 2^i T_i(w_1, w_2, \ldots, w_i) \bmod 2.$$

Remark:[4] T_{2i+1} is divisible by a_1. Hence $w_1 = 0$ implies $U_{2i+1} = 0$.

THEOREM 2.4. *If M^m is almost complex, we have*

$$U^{2i+1} = 0, \qquad U^{2i} = 2^i T_i(c_1, c_2, \ldots, c_i) \bmod 2.$$

The proofs of Theorems 2.1–2.4 are based on the "diagonal" method of Thom[5] and Wu,[4, 6] and on a topological interpretation of the multiplicative sequences which uses the Borel-Serre[7] method of regarding the classes of Stiefel-Whitney, Chern, Pontrjagin as elementary symmetric functions, and which also uses the Whitney duality theorem. The new point in the Theorems 2.1–2.4 lies in the explicit construction of the polynomials and in the fact that for all primes q these polynomials are obtainable from one and the same (rational) multiplicative sequence by reduction mod q.

3. The Index.—For an oriented M^{4k} we can regard the class $L_k(p_1, \ldots, p^k) \in H^{4k}(M^{4k}, R)$ as a rational number. By using a strong theorem of Thom[8] (p. 1735, Theorem 7) we obtain:

THEOREM 3.1. *We have $\tau(M^{4k}) = L_k(p_1, \ldots, p_k)$. Hence L_k for an M^{4k} is always an integer. For example,*

$$3\,\tau(M^4) = p_1, \quad 45\,\tau(M^8) = 7p_2 - p_1^2, \quad 945\,\tau(M^{12}) = 62p_3 - 13p_2p_1 + 2p_1^3.$$

Remarks: It was known to Thom that τ can be expressed as a polynomial in the Pontrjagin classes.[9] Theorem 3.1 implies for an M^4: $p_1 \equiv 0(3)$ (see Wu[6]) and for an M^8: $7p_2 - p_1^2 \equiv 0(45)$. From Theorem 2.1 we obtain $p_1 \equiv 0(3)$ and $7p_2 - p_1^2 \equiv 0(15)$. Analogously for all dimensions. In case M^{4k} is almost complex, we can express the Pontrjagin classes by the Chern classes[3]

$$\sum_{i=0}^{\infty} (-1)^i p_i x^{2i} = \left(\sum_{i=0}^{\infty} c_i x^i \right) \left(\sum_{i=0}^{\infty} (-1)^i c_i x^i \right)$$

and obtain polynomials for τ in the Chern classes.

We now state two theorems about the index which follow from formal properties of the polynomials L_j.

THEOREM 3.2. *If the manifold M^m is fibred in complex projective spaces P_n of n complex dimensions with the group of all projective transformations as structure group and the manifold B^{m-2n} as base ($m \geq 2n$), then*

$$\tau(M^m) = \tau(B^{m-2n})\tau(P_n).$$

Remarks: $\tau(P_n) = 1$, if n is even; $\tau(P_n) = 0$, if n is odd (see the Introduction). For the direct product of two manifolds M, M' we have $\tau(M \times M') = \tau(M) \cdot \tau(M')$.

Consider a manifold M^{4k+2}. Every element $x \in H^2(M^{4k+2}, Z)$ can be represented by a subvariety V^{4k} of M^{4k+2} (Thom[8], p. 573, Theorem 2). The index $\tau(V^{4k})$ only depends on x and may be denoted by $\tau(x)$. If $x_1, \ldots, x_r \in H^2(M^{4k+2}, Z)$, then x_1 can be represented by a subvariety V^{4k} of M^{4k+2}, the restriction of x_2 to V^{4k} can be represented by a subvariety V^{4k-2} of V^{4k}, etc. Finally, the restriction of x_r to $V^{4k-2r+4}$ can be represented by a subvariety $V^{4k-2r+2}$ of $V^{4k-2r+4}$. The index of $V^{4k-2r+2}$ only depends on the non-ordered r-tpl $(x_1, \ldots, x_r)$ and may be denoted by $\tau(x_1, \ldots, x_r)$.

THEOREM 3.3. *We have for $x_1, x_2 \in H^2(M^{4k+2}, Z)$*

$$\tau(x_1 + x_2) = \tau(x_1) + \tau(x_2) - \tau(x_1, x_2, x_1 + x_2).$$

4. The Todd Genus.—Let M_n be an almost complex manifold of n complex dimensions and $c_i \in H^{2i}(M_n, Z)$ its Chern classes. We can regard $T_n(c_1, \ldots, c_n) \in H^{2n}(M_n, R)$ as a rational number which we denote by $T(M_n)$. We call $T(M_n)$ the Todd genus of M_n. Kodaira[10] proved for all algebraic varieties M_n which are a complete non-singular intersection of hypersurfaces in some projective space that

$$T(M_n) = 1 - g_1 + g_2 - \ldots + (-1)^n g_n$$

where g_i is the number of linearly independent i-pl differentials of the first kind attached to M_n. From the results of Todd,[2] Hodge,[11] and Kodaira-Spencer[12] it seems very likely that the last formula is true for all non-singular algebraic varieties. But in the present moment it is not known, even for algebraic varieties, whether $T(M_n)$ is always an integer. Therefore it seems to be interesting that one can prove by the Theorems 2.4 and 3.1:

THEOREM 4.1. *The number* $2^{n-1}T(M_n)$ *is an integer for every almost complex manifold. This means for* $n = 1, 2, 3, 4$:

$$M_1: \ c_1 \equiv 0(2), \quad M_2: \ c_1^2 + c_2 \equiv 0(6), \quad M_3: \ c_1 \cdot c_2 \equiv 0(6),$$

$$M_4: \ -c_4 + c_3c_1 + 3c_2^2 + 4c_2c_1^2 - c_1^4 \equiv 0 \ (90).$$

We state two theorems about the Todd genus which follow from formal properties of the polynomials T_j and which are analogous to the Theorems 3.2 and 3.3.

THEOREM 4.2. *If the almost complex manifold* M_n *is fibred in complex projective spaces* P_k *of* k *complex dimensions with the group of all projective transformations of* P_k *as structure group and the almost complex manifold* B_{n-k} *as base, the fibering being compatible with the almost complex structures of* M_n, P_k *and* B_{n-k} *then*

$$T(M_n) = T(B_{n-k})T(P_k) = T(B_{n-k}).$$

Remarks: $T(P_k) = 1$ for all k. Since $\{T_j\}$ is a multiplicative sequence we have for the direct product of two almost complex manifolds M, M'

$$T(M \times M') = T(M) \cdot T(M').$$

This (in the algebraic case) was checked by Todd[2] for dimensions not exceeding 6.

For a class $x \in H^2(M_n, Z)$ we can define a virtual Todd genus $T(x)$, which is a polynomial of weight n in x and the Chern classes of M_n, such that $T(x)$ is the Todd genus of every admissible almost complex subvariety V_{n-1} of M_n representing x. Moreover, if $x_1, x_2 \in H^2(M_n, Z)$ we can define a virtual Todd genus $T(x_1, x_2)$ with $T(x_1, x_2) = T(x_2, x_1)$ such that for every admissible almost complex subvariety V_{n-1} representing x_1, the number $T(x_1, x_2)$ is the virtual genus with respect to V_{n-1} of the restriction of x_2 to V_{n-1}. The virtual genus $T(x_1, x_2)$ is a polynomial of weight n in x_1, x_2 and the Chern classes of M_n. The following theorem is well known in algebraic geometry:

THEOREM 4.3. *For* $x_1, x_2 \in H^2(M_n, Z)$, *we have*

$$T(x_1 + x_2) = T(x_1) + T(x_2) - T(x_1, x_2).$$

[1] Steenrod, N. E., these PROCEEDINGS, **39,** 213–223 (1953).

[2] Todd, J. A., *Proc. London Math. Soc.*, (2), **43,** 190–225 (1937).

[3] Wu, Wen-Tsun, and Reeb, G., "Sur les espaces fibrés et les variétés feuilletées," *Actual. sci. industr.*, 1183, (1952).

[4] Wu, Wen-Tsun, *Compt. rend. acad. sci., Paris,* **230,** 508–511 (1950).

[5] Thom, R., *Ann. sci. École. norm. sup.* (3), **69,** 109–182 (1952).

[6] Wu, Wen-Tsun, "Sur les puissances de Steenrod," Colloque de Topologie de Strasbourg, 1951; (mimeographed notes).

[7] Borel, A., and Serre, J. P., *Compt. rend. acad. sci., Paris,* **233,** 680–682 (1951); Borel, A., *Ann. Math.*, **57,** 115–207 (1953); Borel, A., and Serre, J. P., *Am. J. Math.*, **75,** 409–448 (1953).

[8] Thom, R., *Compt. rend. acad. sci., Paris*, **236**, 453, 573, 1733 (1953).

[9] Thom, R., "Quelques propriétés globales des variétés différentiables"; (to appear in *Comm. Math. Helv.*).

[10] Kodaira, K., "The Theory of Harmonic Integrals and Their Application to Algebraic Geometry," Notes, Princeton University, 1953.

[11] Hodge, W. V. D., *Proc. London Math. Soc.*, (3), **1**, 138–151 (1951).

[12] Kodaira, K., and Spencer, D. C., "On Arithmetic Genera of Algebraic Varieties," these PROCEEDINGS, **39**, 641–649 (1953).

THE IDEAS OF VARIABLE AND FUNCTION

BY KARL MENGER

ILLINOIS INSTITUTE OF TECHNOLOGY

Communicated by J. von Neumann, June 18, 1953

Investigation of the semantics of pure and applied analysis reveals that there exist three different concepts of a variable, which may conveniently be described as a logical, a scientific, and a mathematical concept. Followers of Weierstrass have treated the logical notion; French analysts, the mathematical and scientific ideas. The confusion of the three concepts accounts for the obscurity which marks the introductions to many treatises on analysis. The lack of a distinction is also one of the reasons that prompted Russell to call the notion of variable "one of the most difficult with which logic has to deal." Severally, the concepts and their mutual relations (which we propose to study within the realm of real numbers) seem clear and simple.

The Logical Concepts of Arithmetized Analysis.—A *function* is a non-empty set of ordered pairs of numbers such that no two pairs of the set contain equal first and unequal second elements. The set of all first (second) elements of the pairs, referred to as arguments (values), is called the domain (the range) of the function. In arithmetized analysis (in conformity with the usage of modern logic), a *real variable* is a symbol which stands for any element of a certain set of real numbers, called the range of the variable. E.g., the set of pairs $(x, \tan x)$, for all numbers x which are not odd multiples of $\pi/2$, is a function which we shall denote by tan. The letter x in the definition of tan and in the formula $D \tan x = \sec^2 x$ is a variable having the domain of tan as range. The symbols log and $\sqrt[3]{}$ denote other functions.

For the set of all pairs (x, x), in spite of the paramount importance of this function, no traditional symbol exists. We shall denote this identity function by ι, and write ι^n, $\mathbf{16} \cdot \iota^2$, and $\mathbf{16}$ for the functions (x, x^n), $(x, 16x^2)$, and $(x, 16)$, respectively, which traditionally are referred to as the functions x, x^n, $16x^2$, and 16, that is, by their values for x. (Even references

to the function tan x are not infrequent, although this symbol for the value is longer than the available symbol, tan, for the function itself.)

Traditionally, $f, g, \ldots$ are used as symbols standing for any function (as "functional variables"). If (x, y) belongs to f, then y is denoted by $f(x)$. For any subset X of the domain of f, we symbolize by $f[X]$ the set of the pairs $(x, f(x))$ for all x belonging to X. Instead of $f[X] = g[X]$ we write $f = g, [X]$. The set of all pairs (x, z) such that, for some y, the pairs (x, y) and (y, z) belong to f and g, respectively, is denoted by gf. This set is a function (i.e., non-empty) if and only if the range of g and the domain of f are non-disjoint. Clearly,

$$\iota f = f = f\iota', \text{ for every } f, \text{ where } \iota' = \iota[\text{dom } f]. \tag{1}$$

gf is called the function obtained by *substitution* of f into g.

In this symbolism, the formulae of classical analysis can be written without numerical variables: $D \tan = \iota^2 \sec$; $D \log = \iota^{-1}$ $[>0]$;

$$\int_0^{\pi} \sin\left(\frac{\mathbf{1}}{\mathbf{2}}\cdot\iota\right) = \int_0^{\pi/2} (\mathbf{2}\cdot\sin) = 2;$$

$$D(f\cdot g) = Df\cdot g + f\cdot Dg, [(\text{dom } Df)\cdot(\text{dom Dg})];$$

$$D(fg) = (Df)g\cdot Dg, [(\text{dom } Dg)\cdot g^{-1}(\text{dom } Df)];$$

$$\int_a^b f = \int_{g^*a}^{g^*b} (fg\cdot Dg). \tag{2}$$

Here, $[>0]$ denotes the set of all positive numbers; $X\cdot Y$ the intersection of X and Y; $g^{-1}X$ the set of all numbers y such that $g(y)$ belongs to X; and g^* the function for which $gg^* = \iota$. The numerical variables a, b in the last formula or in

$$\int_a^b \cos = [\sin]_a^b = \sin b - \sin a \tag{3}$$

can be eliminated if two-place functions $F, G, \ldots$ are defined as certain sets of ordered triples of numbers, e.g., $I, J, H = I^2\cdot J^3$, and $\int \cos$ as the sets of all triples (x, y, x), (x, y, y), (x, y, x^2y^3), and $(x, y, \int^x \cos)$, respectively. Accordingly, for every x and y,

$$I(x, y) = x,\ J(x, y) = y,\ H(x, y) = x^2y^3,\ \int \cos(x, y) = \int^x \cos, \tag{4}$$

and $F(I, J) = F$, $I(F, G) = F$, $J(F, G) = G$, for every F and G if substitution is properly defined. (3) reads $\int \cos = \sin I - \sin J$.

The Scientific Concepts.—The situation in science is altogether different. The variable θ, known as gas temperature, is the pairing with each "specimen" of θ (i.e., with each act of reading a thermometer immersed in gas) of a "value" of θ (i.e., the observed number of degrees). We call the set of all values (specimens) the range (the species) of θ. The gas pressure p is another variable. Suppose that only gas of one and the same volume is under consideration, and that the specimens of θ and p which are simultaneous and concern the same gas are paired. It is a fact of observa-

tion that, if the values for two specimens of p are equal, so are the values for the paired specimens of θ (and vice versa). This is what physicists mean by saying that θ is a function of p (and p a function of θ). If simultaneous specimens of a man's temperature and blood pressure are paired, neither variable is, in general, found to be a function of the other. Each scientific variable, if its specimens are paired with the simultaneous observations of a running clock, is a function of what might be called the historical time. The latter is not, in general, a function of a man's temperature. It is easy to characterize those variables of which time is a function.

By pairing simultaneous specimens of the distance traveled by a falling stone and of the duration of the motion, Galileo discovered not only that s is a function of t, but that $s = 16t^2$. Similarly, $s = \frac{1}{64} v^2$ and $v = 32\ t$. Each such formula may (in analogy to the description of $L = 7^1/_2$ ft. as an equality of a quantity, L, and a "denominate number," $7^1/_2$ ft.) be described as an equality of a variable and a denominate function. We reformulate these equalities in terms of pure numbers and functions by writing (L in ft.) $= 7^1/_2$ and (s sim. t) $= \mathbf{16}\cdot\iota^2$ where (s sim. t) stands for the following function: the set of all ordered pairs (σ, τ) of values of s and t which are paired by virtue of the pairing of the specimens. More briefly, $s_t = \mathbf{16}\cdot\iota^2$, $s_v = \frac{\mathbf{1}}{\mathbf{64}}\cdot\iota^2$, $v_2 = \mathbf{32}\cdot\iota$. The mass of a piece of lead is related to the historical time by a constant function.

A Comparison of the Scientific[1] *and the Pure Concepts.*—Clearly, either of the concepts "pure function" and "scientific variable" can be generalized so as to comprise the other: the latter might be defined as a function with a non-numerical domain, the former as a variable with numerical specimens. However, our distinction reflects the actual use of the terms as far as this use is consistent; moreover, it brings out intrinsic differences between the concepts. A pure function f, in general, permits substitutions of scientific variables and of pure functions—certainly, according to (1) of ι[dom f]. Into a scientific variable neither a variable nor a pure function can be substituted. These differences between the two categories would exist even if both were called functions or both were called variables.

The references to extra-arithmetical specimens and to an extra-arithmetical pairing of specimens—though frequently omitted—are indispensable for sound definitions of scientific variables and functions. Scientific variables are not determined by their range whereas logical variables are.[2] If, for two linear oscillators, the positions at various instants t are found to be sin t and cos t, respectively, then the second variable is not a function of the first if simultaneous specimens are paired; but it is if, for each t,

the specimen at t of the first variable is paired with that of the second at $2t$.

The Mathematical Concepts of a Non-Arithmetized Analysis.—Between the theories of pure functions (sin, $16 \cdot \iota^2$, . . .) and of functional connections (s_t, s_v, v_t, . . .) between scientific variables, there is a mathematical theory dealing with undefined entities (comparable to postulational geometry of points and mappings) which describes, as it were, the applications of pure functions to science. A *mathematical variable* u is a set of pairs (a, α) in which a is an element of an abstract set A (the species of u) and α belongs to a set U of numbers (the range of u). By a *pairing* of the species A and B we mean a set P of pairs (a, b) such that each element of A, as well as each element of B, belongs to exactly one pair of P. The variable v with the species B is a *mathematical function* of u with respect to P if the set of pairs (α, β) for all pairs (a, b) of P is a pure function. We denote the latter by $(v\ P\ u)$ or, if only one pairing of A and B is considered, briefly, by v_u. In this function, u is called the independent, v the dependent variable.

The main statements of the theory (which can be axiomatized without reference to numbers) include: $(u \text{ id. } u) = \iota[U]$, where id. is the set of all pairs (a, a). If $(w\ P\ v) = f$ and $(v\ Q\ u) = g$, then $(w\ PQ\ u) = fg$ (substitution of g into f). If $v_u = g$ and $z_u = h$ and h^* exists, then $v_z = gh^*$—the substitutive contravariance of the dependent to the independent variable.[3] Here $hh^* = \iota$.

Calculus is connected with the mathematical theory of variables by the assumption that for many pairs of variables for which $(w\ P\ v) = w_v$ exists, there is a variable, denoted by dw/dv and called the rate of change of w with v, having the same species as v and connected with v (with regard to id.) by the function $(dw/dv)_v = D\ w_v$. From the preceding theorems and (2) we readily derive the basic equality

$$(dw/dv)_u \cdot (dv/du)_u = (dw/du)_u.$$

If $v_u = \cos$, then $\int_a^b v_u$ (which traditionally is written $\int_a^b v\ du$), by (3), equals $[\sin]_a^b$. This result can, in turn, be applied to scientific variables. If an elastic force p is related to the time as v is to u above (that is to say, if $p_t = \cos$), then $\int_a^b p_t = [\sin]_a^b$. The last integral (written $\int_a^b p\ dt$) is the impulse. However, results about scientific variables can also be obtained directly from formulae concerning pure functions. E.g., if p is related to the position s of the oscillator by Hooke's function $p_s = -\iota$, then the work is $\int_a^b p_s = \int_a^b (-\iota) = \left[-\frac{1}{2} \cdot \iota^2\right]_a^b$. It should further be noted that, in contrast to $\int_b^a p_t$ and $D\ p_s$, the symbols $\int_a^b p$ and $D\ p$ are meaningless, since p is a variable while integrals and derivatives are defined for functions only.

Semantic Inadequacies of Traditional Analysis.—The confusion of the

three types of variables with each other and with functions is partly due to an unfortunate notation. The founders of analysis denoted (1) the values for x and 3 of a pure function f (such as sin or $\mathbf{16}\cdot\iota^2$) by $f(x)$ and $f(3)$; (2) the function connecting the scientific variables s and t by $s(t)$ (and they expressed the fact that s is a function of t by $s = s(t)$); (3) the logical variable in formulae such as $D \tan x = \sec^2 x$ by x, and a mathematical variable in statements such as "the variable y is a function of the variable x" by the same letter x.

A comparison of the symbols $s(t)$ and $f(x)$ suggests that s is a function like f or sin, and that x is a variable of the same type as t. But neither is the case, as evinced by the fact that, as x is replaced by 3 in $f(x)$ or sin x, the result is a definite number, $f(3)$ or sin 3, whereas replacement by 3 of t in $s(t)$ results in the incomplete symbol $s(3)$, eschewed by physicists because it might be interpreted in many ways, e.g., as the value for 3 of s_v. It is the connection of s with t, rather than s, which corresponds to a pure function. In fact s_t is *equal* to the pure function $\mathbf{16}\cdot\iota^2$. The analog of $f(3)$ or sin 3 is the value $s_3(3) = 16\,\iota^2 3 = 144$.

In applications of y_x to science, the mathematical variables x and y may be replaced by scientific variables, such as s, t, v. They must not be replaced by numbers while the logical variable x in sin x or in $f(x)$ may. These remarks clarify the traditional use—bordering on contradiction—of $s(v)$ beside $s(t)$, of $s = s(v)$ beside $s = s(t)$, and of such formulae as $u = u(p, \theta) = u(p, v) = u(v, S) = \ldots$. Physicists realize that they are dealing with the same variable, S or u, in different connections, and therefore prefer the above notation to $s_1(t)$ and $s_2(v)$, etc. However, the subscripts would be necessary if s and u stood for functions. Actually, the same scientific variable s is connected with t by the function $s_t = \mathbf{16}\cdot\iota^2$ and with v by the function $s_v = \frac{\mathbf{1}}{\mathbf{64}}\cdot\iota^2$. Similarly, u is connected with different pairs of scientific variables by different two-place (pure) functions, $u_{p\theta}$, u_{pv}, u_{vS} . . . . Their values for (a, b) can be denoted by $u_{p\theta}(a, b)$, etc., their partial derivatives by $D_1 u_{p\theta}$, $D_2 u_{p\theta}$, $D_1 u_{pv}$.

[1] In *geometry*, because of the equivocal use of this term, the situation is not uniform. In the *analytic* Cartesian plane, each point is defined as an ordered pair of numbers. Hence, according to (4), the abscissa is the two-place (pure) function I. In a plane of *observation* with a Cartesian frame of reference, the abscissa is a scientific variable. In a plane of *postulational* geometry, points and lines are undefined entities or elements of abstract sets. If three non-collinear points, 0, 1_x, 1_y, are selected, then the abscissa is what might be called a "geometric variable"; namely, the pairing (by construction) of a numerical value with each punctual specimen. A similar situation exists for the radius r and the volume v of a sphere, where $(v\,\mathrm{id}\cdot r) = \frac{4}{3}\pi s^3$, and in differential geometry.

[2] p and θ have the same range $[>0]$ and yet cannot always be substituted for one another whereas logical variables x and y with the same range can. E.g., the formulae

$D \log x = \iota^{-1}x$ and $D \log y = \iota^{-1}y$ have the same meaning, namely, $D \log = \iota^{-1}[>0]$. Similarly, both $D_1(I^2 \cdot J^3)(x, y) = 2xy^3$ and $D_1(I^2 \cdot J^3(y, x) = 2yx^3$ stand for $D_1(I^2 \cdot J^3) = 2 \cdot I \cdot J^3$. Here, the issue is somewhat confused by the traditional reference to $I^2 \cdot J^3$ as the function x^2y^3, that is, by its value for (x, y). The connection of the alphabetical order of the variables with the places where they are to be substituted is not germane to the problem.

[5] The above theory is strikingly analogous to a general theory of mensuration dealing with undefined objects A, B, . . . and numbers such as $(B \text{ in } A)$. In this theory, $(A \text{ in } A) = 1$; if $(B \text{ in } A) = b \neq 0$, then $(A \text{ in } B) = 1/b$; if $(B \text{ in } A) = b$ and $(C \text{ in } B) = c$, then $(C \text{ in } A) = c \cdot b$; and measure is multiplicatively contravariant to a change of the unit.

PAIRS OF MATRICES WITH PROPERTY L

BY T. S. MOTZKIN AND OLGA TAUSSKY

UNIVERSITY OF CALIFORNIA AT LOS ANGELES AND NATIONAL BUREAU OF STANDARDS

Communicated by J. L. Walsh, June 12, 1953

This note is concerned, for matrices with elements in an algebraically closed field of characteristic zero, with pencils generated by pairs of matrices with property L.

A pair of $n \times n$ matrices is said to have property L if, for a special ordering of the characteristic roots α_i of A and β_i of B, the characteristic roots of $\lambda A + \mu B$ are $\lambda\alpha_i + \mu\beta_i$ for all values of λ and μ (see ref. 1). Another characterization of pairs of matrices with property L can be given for a large class of such pairs. It is obtained by considering those matrices in the pencil $\lambda A + B$ which have multiple characteristic roots. Unless every matrix in the pencil has a multiple characteristic root there are at most $n(n-1)$ matrices which do. This follows easily from the facts that the equation $|\lambda A + B - \nu I| = 0$ has multiple roots ν if and only if the discriminant Δ of the polynomial in ν on the left vanishes and that Δ is a polynomial in λ of degree $n(n-1)$ or less. Call the values of λ (including ∞) for which $\Delta = 0$ the *discriminant roots* of the pencil.

Definition: A pair of matrices A, B has property D if $\Delta \equiv 0$ or if all discriminant roots are of even order.

It can be shown that a pair of matrices with property L has property D. A partial converse is:

THEOREM 1. *Let* A, B *be a pair of matrices with property* D. *Assume that no matrix* λA $+ \mu$B (*for* $\lambda, \mu \neq 0, 0$) *has a characteristic root of multiplicity* $\geqq 3$ *nor two different double characteristic roots. It then follows that* A, B *have property* L.

The condition that no matrix in the pencil has a triple characteristic root is necessary as is shown by the following example:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The matrices A, B, $A - B$ all have a triple characteristic root. A, B do not have property L since A has all its characteristic roots zero, B is singular, while $A + B$ is not singular. All the discriminant roots are double.

The methods used for the proof of this theorem enable us to prove the following further theorems which deal with general pencils of matrices.

THEOREM 2. *Let* $\lambda A + \mu B$ *be a pencil in which all matrices are diagonable, i.e. similar to diagonal matrices. Then* A, B *have property* L.

However, an even stronger result can be established.

THEOREM 3. *Let* $\lambda A + \mu B$ *be a pencil in which all matrices are diagonable. It follows that* A, B *commute.*

Finally, it can be shown that any pair of commutative matrices can—in the case of complex numbers—be regarded as the limit of a pair of diagonable matrices. For elements in a general field we obtain the following theorem:

THEOREM 4. *Every generic pair of commutative matrices is diagonable.*

In order to prove Theorems 1 and 2 the following link with the theory of algebraic curves is introduced. The characteristic roots of a matrix $\lambda A + \mu B$ are the solutions ν of the determinantal equation

$$|\lambda A + \mu B - \nu I| = 0$$

for fixed λ, μ. Interpret this equation as the equation of an algebraic curve C in the projective λ, μ, ν-plane. We call C the *characteristic curve* of the pencil. To say that A, B have property L is equivalent to saying that the curve C splits into straight lines.

The point $P = (0, 0, 1)$ is never on the curve C. Consider a tangent t from P to C and let T be one of its points of contact. Then the matrix $\lambda_0 A + \mu_0 B$ which corresponds to T has a multiple characteristic root. For, let (λ_0, μ_0) be an arbitrary pair of numbers and $\nu_1, \ldots, \nu_n$ the characteristic roots of $\lambda_0 A + \mu_0 B$. The n points $(\lambda_0, \mu_0, \nu_i)$ are all collinear with P. If two of these points coincide, i.e., if two of the ν_i coincide this line will be a tangent to C. If more than two of the ν_i coincide, i.e., if there is at least a triple characteristic root, then T will be a singular point of C or an inflection point.

The correspondence thus set up between the matrices in the pencil with multiple characteristic roots and the tangents from P to the curve C is also a correspondence between these tangents and the discriminant roots. In particular a multiple discriminant root corresponds to a tangent with a singular or inflection point as point of contact or with at least two points

of contact. If every matrix $\lambda_0 A + \mu_0 B$ has a multiple characteristic root it follows that C has a component which has to be counted double.

Detailed proofs and extensions to fields of finite characteristic will appear elsewhere.

[1] Motzkin, T. S., and Taussky, Olga, "Pairs of Matrices with Property L," *Trans. Amer. Math. Soc.*, **73**, 108–114, (1952).

ON RESTRICTED PARTITIONS AND A GENERALIZATION OF THE EULER φ NUMBER AND THE MOEBIUS FUNCTION

BY CHARLES A. NICOL

UNIVERSITY OF TEXAS

Communicated By H. S. Vandiver, June 23, 1953

Introduction.—In the present paper we shall treat the function

$$F_{n-1}(z, x) = \prod_{s=1}^{n-1} (z - x^s) \tag{1}$$

mainly from an arithmetic standpoint.

Functions of this type have been studied extensively in the theory of partitions of positive integers[1] where z is replaced by 1 or -1 and the range of the product is infinite. For example, a famous result due to Euler for $|x| < 1$ may be written as

$$(1 - x)(1 - x^2)(1 - x^3) \ldots = 1 + \sum_{n=-\infty}^{\infty} (-1)^n x^{(1/2)n(3n+1)}. \tag{2}$$

The coefficients of the series admit the following combinatorial interpretation. If $E(n)$ denotes the number of partitions of n into an even number of unequal parts and $U(n)$ the number of partitions of n into an odd number of unequal parts then (2) may be stated in the following way.

$E(n) = U(n)$ except when $n = \frac{1}{2}k(3k \pm 1)$, when $E(n) - U(n) = (-1)^k$

The case where we have only a finite number of terms in (2) has received comparatively little attention. If in (1) we let $z = 1$ we have

$$F_{n-1}(1, x) = \prod_{s=1}^{n-1} (1 - x^s). \tag{3}$$

The coefficient of x^k for $k \leq n$ will be $E(k) - U(k)$. But the coefficient of x^k for $k > n$ is the number of partitions of k into an even number of unequal parts, none of which is larger than n, minus the number of partitions of k

into an odd number of unequal parts, none of which is larger than n. This then is an extension of the analogous problem arising in the use of the infinite product.

Also if z is replaced by -1 we consider the following function.

$$(-1)^{n-1}F_{n-1}(-1, x) = \prod_{s=1}^{n-1}(1 + x^s). \tag{4}$$

The coefficient of x^k resulting from the expansion of this product is the number of partitions of k as a sum of distinct positive integers none of which is larger than n. This may also be stated as the number of solutions of the equation $x_1 + 2x_2 + \ldots + (n-1)x_{n-1} = k$, where for $i = 1, 2, \ldots, (n-1)$, x_i is either zero or unity.

The products (3) and (4) have been studied by Cauchy, T. Vahlen, von Sterneck, and others.[2] In particular von Sterneck studied the case where the polynomial resulting from the expansion is reduced modulo a positive integer.

Fundamental in this investigation will be the use of the number[3]

$$\Phi(k, n) = \frac{\varphi(n)}{\varphi(n/(k, n))}\mu(n/(k, n)), \tag{5}$$

where k and n are positive integers and (k, n) denotes the greatest common divisor of k and n. If n is a positive integer $\varphi(n)$ denotes as usual the number of positive integers less than n and prime to it. ($\varphi(1) = 1$.) Also, for n a positive integer, $\mu(n)$ is zero if n contains a repeated prime factor. Otherwise $\mu(n)$ is equal to $(-1)^\gamma$ where γ is the number of distinct prime factors of n. ($\mu(1) = 1$.) Note that (5) reduces to $\mu(n)$ when $(k, n) = 1$ and $\varphi(n)$ when $(k, n) = n$. Although (5) appears more complicated than its constituents it will be shown that many of the principal theorems concerning it are hardly more complex than those involving the φ or μ number alone.

The properties of the coefficients in the development of (1) are extensively used to obtain properties of (5) and vice versa.

We now state a number of theorems without proof. We hope to publish the proofs elsewhere.

In the following paragraphs the symbol $[x]$ will denote the largest integer contained in the real number x. Also the symbol $\Phi(a, b)$ will denote the number defined in (5) where a and b are positive integers.

We have if k, r, and n are positive integers that[4]

$$\sum_{(r,n)=1} \exp(2\pi irk/n) = \Phi(k, n), \tag{6}$$

where the range of r is over all positive integers less than n and prime to it. ($i^2 = -1$.)

We also have

THEOREM 1.

$$\sum_{d|n} \Phi(k, d) = \begin{cases} n \text{ if } (k, n) = n. \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

Similarly

$$\sum_{d|n} (-1)^d \Phi(k, n/d) = \begin{cases} 0 \text{ if } (k, n) \neq \text{n.} \\ -n \text{ if } (k, n) = n, n \text{ even.} \\ 0 \text{ if } (k, n) = n, n \text{ odd.} \end{cases} \tag{8}$$

We may also prove

THEOREM 2. *Let $\sigma_n(k)$ denote the sum of the divisors of k less than or equal to n. Then,*

$$\sum_{s=1}^{n} [n/s]\Phi(k, s) = \sigma_n(k). \tag{9}$$

In case $k = 1$ this becomes the well-known relation[5]

$$\sum_{s=1}^{n} [n/s]\mu(s) = 1.$$

Also if k is replaced by $n!$ we have another known result[6]

$$\sum_{s=1}^{n} [n/s]\, \varphi(s) = n(n + 1)/2.$$

THEOREM 3. *If $\delta \mid n$, then*

$$\sum_{d|\delta} \binom{dn/\delta}{d} (-1)^d \Phi(k, \delta/d) \equiv 0 (\text{mod. } n). \tag{10}$$

COROLLARY. *If now p denotes an odd prime and α is a positive integer, then*

$$\binom{n}{p^\alpha} \equiv \binom{n/p}{p^{\alpha-1}} (\text{mod. } n). \tag{10a}$$

THEOREM 4.

$$\sum_{d|n} R_s(d)\Phi(k, n/d) \equiv 0 \text{ (mod. } n), \tag{11}$$

where

$$R_s(d) = \sum_{\alpha} \binom{d-1}{\alpha} (-1)^\alpha,$$

and this sum is over all integral solutions α of the equation $(n/d)\alpha + \beta = s$ where $0 \leq s < n, 0 \leq \alpha < d, 0 \leq \beta < n/d$.

Consider now the function $F_{n-1}(z, x)$ defined in (1). If $n > 1$ and the product is expanded as a polynomial in x, we may write

$$F_{n-1}(z, x) = \sum_{s=0}^{\bar{n}} P_s(z)x^s, \tag{12}$$

where $\bar{n} = n(n-1)/2$ and $P_s(z)$ is a polynomial in z. Then we may define the polynomial $B_t(z)$ as

$$B_t(z) = \sum_{k=0}^{n_t} P_{kn+t}(z), \tag{13}$$

where $n_t = [(n-1)/2 - t/n]$ and $1 \leq t \leq n$.

Then we may obtain

THEOREM 5. *If z is a number different from unity, then*

$$B_t(z) = \frac{1}{n(z-1)} \sum_{d|n} (z^{n/d} - 1)^d \Phi(t, n/d), \tag{14}$$

where $1 \leq t \leq n$ and $B_t(z)$ is defined in (13).

Denote the polynomial defined by the function $F_{n-1}(1, x)$ in (3) by

$$\sum_{s=0}^{n} A_s x^s, \text{ where } \bar{n} = n(n-1)/2. \tag{15}$$

Also define the number C_t by the relation

$$C_t = \sum_{k=0}^{n_t} A_{kn+t}, \text{ where } n_t = [(n-1)/2 - t/n]. \tag{15a}$$

Then we obtain the following

THEOREM 6.

$$C_t = \Phi(t, n), \tag{16}$$

where C_t is defined in (15a).

THEOREM 7.

$$\varphi(n) = \frac{1}{n} \sum_{t=1}^{n} C_t^2, \tag{17}$$

where C_t is defined in (15a).

In view of theorem 6 we may write theorem 7 as

$$\varphi(n) = \frac{1}{n} \sum_{t=1}^{n} \Phi^2(t, n). \tag{18}$$

An observation of possible interest may be made concerning theorem 7 if it is noted that $C_n = \varphi(n)$. Then (17) becomes a quadratic relation in $\varphi(n)$. Employing the quadratic formula we find

$$\varphi(n) = (n \pm \sqrt{n^2 - 4G(n)})/2, \tag{19}$$

where $G(n) = \sum_{t=1}^{n-1} C_t^2$.

Except in the case when $\varphi(n) = n/2$ only one of the roots of (19) corresponds to $\varphi(n)$. The significance of the remaining root has not been determined and would seem to be of interest.

If x is replaced by $\exp(\iota\theta)$, where $\iota^2 = -1$, we obtain the following

THEOREM 8.

$$\Phi(t, n) = \frac{1}{2\pi} \int_0^{2\pi} \left\{ F_{n-1}(\exp(i\theta)) \sum_{k=0}^{n_t} \exp(-(kn + t)i\theta) \right\} d\theta, \tag{20}$$

where $F_{n-1}(exp(i\theta)) = \prod_{s=1}^{n-1} (1 - exp(si\theta))$ *and* $n_t = [(n - 1)/2 - t/n]$. *Similarly for the numbers* A_t *defined in (15) we have*

$$A_t = \frac{1}{2\pi} \int_0^{2\pi} \{F_{n-1}(\exp(i\theta))\ (\exp(-ti\theta))\}\, d\theta. \tag{21}$$

Furthermore the numbers A_t defined in (15) have the following properties:

$$\sum_{s=0}^{\bar{n}} |A_s| \geq n, \tag{22}$$

where $\bar{n} = n\ (n - 1)/2$.

A rather unusual property of these numbers is:

THEOREM 9. *If* $d|(n - 1)$, *then*

$$\sum_{(s,\ n-1)=d} A_s = 0, \tag{23}$$

where $0 \leq s \leq n(n - 1)/2$.

A by-product of these investigations is the following result: If p is an odd prime, the integral roots of the congruence

$$1 + \sum_{s=1}^{p-1} \Phi(s, p - 1)x^s \equiv 0 \pmod{p} \tag{24}$$

are the incongruent primitive roots modulo p.

If we consider the function defined in (4) we may obtain the following: If n is an odd positive integer and t is an integer such that $1 \leq t \leq n$, then

$$B_t(-1) = \frac{1}{2n} \sum_{d|n} 2^d\, \Phi(t, n/d), \tag{25}$$

where $B_t(z)$ is defined in (13).

Since $B_t(z)$ is a polynomial with integral coefficients the number $B_t(-1)$ is an integer.

In particular, if n is odd and t is less than n and prime to it, then

$$B_t(-1) = \frac{1}{2n} \sum_{d|n} 2^d \mu(n/d). \tag{26}$$

This number is a generalization of the Fermat Quotient, $(2^{p-1} - 1)/p$, where p denotes an odd prime.

Acknowledgment.—The author is indebted to H. S. Vandiver for his many helpful suggestions and encouragement.

[1] Dickson, L. E., *History of the Theory of Numbers*, Vol. 2, chapter 3, Carnegie Inst. of Washington, Publication No. 256 (1920).

Bachmann, P., "Niedere Zahlentheorie," Zweiter Teil, chapters 3–6, B. G. Teubner, Leipzig, 1910.

[2] Vahlen, T., Bachmann, P., *Ibid.*, 116, 167, 273.

Von Sterneck, *Sitzungsber. d. Wiener Akad.*, **111**, 1567 (1902); **113**, 326 (1904); **114**, 711 (1905).

Cauchy, *Oeuvres D'Augustin Cauchy*, **5** (series 1), 81–85, 135–152. Paris, Gauthier-Villars (1885).

[3] Von Sterneck (P. Bachmann, *Ibid.*) introduced a function equivalent to $\Phi(k, n)$. He used it to obtain results concerning partitions modulo a positive integer. Employing this function he obtained a special case of theorem 1.

The number $\Phi(k, n)$ was used by R. Moller in the following result (*Math. Monthly*, **59**, No. 4, 228 (April 1952)). If the numbers g_d are all of the incongruent integers belonging to d modulo p, p being an odd prime and d a divisor of $p - 1$, then for any r, $\sum g_d^r \equiv \Phi(r, d) \pmod{p}$

[4] The sum $\sum_{(r,n)=1} \exp(2\pi irk/n)$ is known as Ramanujan's sum (cf., Hardy, G. H. and Wright, E. M., *Introduction to the Number Theory*, Oxford, 1938. pp. 55, 237). Another closed form for this sum was found previously by T. M. Apostal and D. R. Anderson and stated by them in an abstract in *Bull. Am. Math. Soc.*, **58**, No. 5, 559 (1952). The form they found in our notation is $c\mu(b)\varphi(a)/\varphi(c)$ where $a = (n, k)$: $b = n/a$, and $c = (a, b)$. If $\mu(b) \neq 0$ we have the relation $\varphi(n)/\varphi(b) = c\varphi(a)/\varphi(c)$.

[5] Nagell, T., *Introduction to Number Theory*, Uppsala, 1951, p. 43.

[6] Review of Perez-Cacho, "The Function $E(x)$ in the Theory of Numbers," *Math. Reviews*, **18**, 913, (1952).

CHARACTERIZATION OF QUADRATIC FORMS OVER LOCAL FIELDS

BY O. T. O'MEARA

MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY

Communicated by S. Lefschetz, June 29, 1953

1. *Local Fields.*—Let F be a field with a complete, discrete valuation, and having a finite residue class field. Then F is said to be a local field. Local fields are characterized as (1) the finite extensions of the p-adic numbers and (2) the fields of formal power series over finite constant fields. In particular, the p-adic numbers constitute a local field.

In the proofs of our results, it will be found necessary to perform divisions by the number 2; therefore, assume in addition that F does not have characteristic 2.

Define $e = \text{ord } 2$ as the ramification index of F. If π denotes a prime in F, $2/\pi^e$ is a unit. For example, $e = 0$ if F is the p-adic numbers, $p \neq 2$. The 2-adic numbers have ramification index 1.

2. *Quadratic Forms.*—Consider the non-degenerate quadratic form $f = \sum^{n} a_{ij}x_ix_j$, with $a_{ij} = a_{ji}$, defined over F. Then $g = \sum^{m} b_{ij}y_iy_j$ is rationally equivalent to f—i.e., by means of a rational linear transformation—if and only if: (1) $n = m$; (2) their determinants differ by a square factor: i.e., $d(f)/d(g)$ is a square; (3) their Hasse symbols are the same: $S(f) = S(g)$. However, we are more concerned with integral equivalence in this paper; that is, under unimodular integral transformations.

There will be no loss of generality in restricting ourselves to integral forms only—i.e., a_{ij} are integers in F. The matrix $\mathfrak{F} = (a_{ij})$ will be called the matrix of f.

3. *The Canonical Representation of f.*—Let f be an integral form, the determinant of whose matrix is a unit. Then (1) if $e = 0$, f is integrally equivalent to a diagonal form; (2) if $e \geq 1$, either f can be diagonalized and is then said to be proper, or, if not, it can be expressed as a sum of forms of the type $\alpha x^2 + 2xy + \beta y^2$, where α and β have ordinal greater than zero.

If f is a general integral form, then unimodular transformations can be found, sending f into $f^* = \sum_{1}^{p} \pi^{s_i} f_i$ where the f_i are integral forms of unit determinant and $s_{i+1} > s_i$.

If $g' = \sum_{1}^{q} \pi^{t_i} g_i$, we say that f and g have the same type if (1) $p = q$; (2) $t_i = s_i$ all $i \leq p$; (3) g_i and f_i have the same number of variables for each $i \leq p$; (4) f_i and g_i are proper or improper together for each $i \leq p$.

Two canonical representations f^* and f^{**} of f have the same type.

The quantity s_p is the order of f.

4. *Complete Set of Invariants When $e = 0$.*—f is integrally equivalent to g if and only if they are of the same type and $d(f_i)/d(g_i)$ is a square for each $i \leq p$.

It is convenient to have a characterization independent of the canonical form. Define α_i as the least value of the ordinals of all i by i minors of the matrix $\mathfrak{F}$. Let e_i be the ratio α_i/α_{i-1} with $e_1 = \alpha_1$. If $\Phi(i)$ denotes the matrix of some $i \times i$ principal minor whose determinant has least ordinal, and if $\phi(i) = \text{ord } (d\Phi(i))$, then f is characterized by the quantities (1) e_i, $1 \leq i \leq n$. (2) $(d\Phi(i)/\pi)$ for $1 \leq i \leq n$, i satisfying the condition $e_i \neq e_{i+1}$. $[(d\Phi/\pi^k)$ is defined so that $(a/\pi^k) = (b/\pi^k)$ if and only if a/b is both a unit and a square modulo π^k.]

5. *The Unramified Case, $e = 1$.*—If f has the canonical representation $\sum_1^r \pi^{s_i} f_i$, then f is said to be a continuous form if

(1) $s_{i+1} = s_i + 1$ for $1 \leq i \leq r - 1$.

(2) f_i is proper for $1 < i \leq r$ and proper or improper for $i = 1$.

Then f can be expressed as the sum of maximal continuous forms: $\sum_1^q M_j = \sum m_{\lambda\mu} x_\lambda x_\mu$. Assume that this is still a canonical representation of f. Let $i(j)$ be the number of variables in $M_1 + \ldots + M_j$.

$i(j) \to i(j) + 1$ is said to be a discontinuity of the first type if $\Delta[i(j)] = [m_{i(j)-1} x_{i(j)-1}^2 + 2m_{i(j),\,i(j)-1}\, x_{i(j)} x_{i(j)-1} + m_{i(j)} x_{i(j)}^2] + [m_{i(j)+1} \times x_{i(j)+1}^2 + \ldots]$ has one of the following forms:

$$(1) \qquad \pi^t \begin{pmatrix} \alpha & 1 \\ 1 & \beta \end{pmatrix} + \pi^{t+s}(\epsilon) \ldots \qquad s \geq 2.$$

$$(2) \qquad \pi^t \begin{pmatrix} \alpha & 1 \\ 1 & \beta \end{pmatrix} + \pi^{t+s} \begin{pmatrix} \alpha' & 1 \\ 1 & \beta' \end{pmatrix} \qquad s \geq 1.$$

$$(3) \qquad \ldots\ldots \pi^t(\delta) + \pi^{t+s}(\epsilon) \ldots \qquad s \geq 3.$$

$$(4) \qquad \ldots\ldots \pi^t(\delta) + \pi^{t+s} \begin{pmatrix} \alpha' & 1 \\ 1 & \beta' \end{pmatrix} \qquad s \geq 2.$$

$$(5) \qquad i = n.$$

This discontinuity is of the second if $\Delta[i(j)]$ is one of:

$$(6) \qquad \ldots\ldots \pi^t(\delta) + \pi^{t+2}(\epsilon) \ldots.$$

$$(7) \qquad \ldots\ldots \pi^t(\delta) + \pi^{t+1} \begin{pmatrix} \alpha' & 1 \\ 1 & \beta' \end{pmatrix}.$$

ϵ and δ are units. $\alpha, \alpha', \beta, \beta'$ denote elements of ordinal ≥ 1.

The integral equivalence class of f is then characterized by

(1) the type of f.

(2) $\left(\frac{d(M_1 + \ldots + M_j)}{\pi^{v_j}}\right)$ where $v_j = \begin{cases} 2 \text{ if dis. of 2nd type at } j. \\ 3 \text{ if dis. of 1st type at } j. \\ 0 < j \leq q. \end{cases}$

(3) $(\pi^{\text{ord}[\,d(\sum_1^j M_i)\,]\,+\,m_j},\ d(\sum_1^j M_i)) \cdot S(\sum_1^j M_i)$ for all j, $0 < j \leq q$, m_j being the order of M_j.

Once again we express the invariants in a form which does not depend on a canonical expression of the quadratic form. For each i, $1 \leq i \leq n$, let $\Phi(i)$ be any principal $i \times i$ minor of $\mathfrak{F}$ whose determinant has minimal ordinal. $\phi(i) = \text{ord}\,[d(\Phi(i))]$. Let T be the set of integers defined by

$$T = \{i \mid 0 < i \leq n,\ \phi(i+1) + \phi(i-1) - 2\phi(i) \geq 0\}.$$

If $i \epsilon T$, put

$$h(i) = \frac{\phi(i+1) - \phi(i-1)}{2} - 1$$

If $i \epsilon T$, put

$$h(i) = \begin{cases} \phi(i+1) - \phi(i), \text{ if } i+1 \epsilon T \\ \dfrac{\phi(i+2) - \phi(i)}{2} + 1, \text{ if } i+1 \epsilon T. \end{cases}$$

And $h(-1) = -\infty$, $h(n) = \infty$. This defines $h(i)$ for $-1 \leq i \leq n$.

THEOREM. *f is characterized by*

(1) $h(i)$ when $0 \leq i \leq n-1$.

(2) For values of i satisfying $1 \leq i \leq n$; $h(i) > h(i-2)$; $h(i) - h(i-1) \geq 2$; the quantities (a) $(d\Phi(i)/\pi^{h(i) - h(i-1)})$; (b) $(\pi^{\phi(i) + h(i-1)}, d\Phi(i)) \times S(\Phi(i))$.

The Case $e > 1$.—A complete set of invariants is not known when 2 splits in F. However, the problem of integral equivalence can be reduced to a finite one by the

THEOREM. *Let f and g have the same order m, and let their matrices be $\mathfrak{F}$ and G. Then f and g are equivalent if $\mathfrak{F} \equiv \text{G} (\text{mod } \pi^{2e+m+1})$.*

In the cases $e \leq 1$, it is easy to see that two integral forms of unit determinant and of the same type are integrally equivalent if they are rationally equivalent. If $e > 1$, this is not true. However, two rationally equivalent forms of unit determinant are integrally equivalent if they represent the same numbers as modulo π^{2e+1}. The proof of this is not easy.

THEOREM. *If f and g are integral forms of unit determinant, then f and g are integrally equivalent if and only if (1) they have the same number of variables, (2) they represent the same numbers modulo* π^{2e+1}, *(3)* $S(f) = S(g)$, *and (4)* $d(f)/d(g)$ *is a square.*

THE AXIOM OF CHOICE IN QUINE'S NEW FOUNDATIONS FOR MATHEMATICAL LOGIC

BY ERNST P. SPECKER

EIDGEN. TECHNISCHE HOCHSCHULE, ZURICH

Communicated by H. Weyl, July 3, 1953

The object of this note is to disprove the axiom of choice in Quine's "New Foundations."[1] As the axiom of choice is provable for finite sets, we obtain the axiom of infinity as a corollary.[2]

There will be references to Rosser's *Logic for Mathematicians*,[3] although the axiom of infinity is assumed there; but it is readily eliminated for our purposes if one replaces Quine's ordered pair by Kuratowski's.

The proof will be by *reductio ad absurdum*.

1.1 V is the universal set, Λ the null set (R 256).

1.2 $SC(a)$ is the set of subsets of a (R 255).

1.3 $SC(V) = V$ (R 256).

1.4 $USC(a)$ is the set of unit subsets of a, $USC^2(a) = USC(USC(a))$ (R 255).

1.5 *Fin* is the set of finite sets (R 417).

2.1 Cardinal numbers are construed as saturated sets of equivalent sets (R 371). $Nc(a)$ is the cardinal number of a; so $a \in Nc(a)$.

2.2 NC is the set of cardinal numbers; $\Lambda \notin NC$.

2.3 $Nc(SC(a)) = Nc(SC(b))$ if $Nc(a) = Nc(b)$ (R 369).

2.4 $Nc(USC(a)) = Nc(USC(b))$ if and only if $Nc(a) = Nc(b)$ (R 368).

2.5 $Nc(SC(V)) = Nc(V)$ (1.3).

2.6 $Nc(SC(USC(a))) = Nc(USC(SC(a)))$ (R 368).

2.7 $Nc(SC(USC(V))) = Nc(USC(V))$ (2.6 and 1.3).

2.8 A cardinal number is finite if it is a subset of *Fin*. *FNC* is the set of finite cardinal numbers.

2.9 1, 2, 3 are defined as cardinal numbers of sets with one, two, three elements; $1, 2, 3 \in FNC$.

3.1 Definition of the sum of two cardinal numbers m, n: $m + n = Nc(a \cup b)$ if $m = Nc(a)$, $n = Nc(b)$ and $a \cap b = \Lambda$; if there are no such a, b, then $m + n = \Lambda$ (R 373).

3.2 $1 + 1 = 2$, $1 + 2 = 3$.

3.3 If n is a finite cardinal number and $n + 1 \neq \Lambda$, then $n + 1$ is a finite cardinal number.

3.4 If m is a finite cardinal number, then there are finite cardinal numbers n, p, q such that either $m = n + n + n$ or $m = p + p + p + 1$ or $m = q + q + q + 2$; either of these three cases excludes the two others.

3.5 The cardinal numbers are well-ordered by the relation "there are sets a, b such that $a \in m$, $b \in n$ and $a \subseteq b$" (axiom of choice).

3.6 If $n + 1$ is a finite cardinal number, then $n < n + 1$.

4.1 Definition of 2^m for cardinal numbers m: If $m = Nc(USC(a))$, then $2^m = Nc(SC(a))$ (cf. 2.3, 2.4; R 389); if there is no set a such that $USC(a) \in m$, then $2^m = \Lambda$.

4.2 $2^{Nc(USC)(a))} = Nc(SC(a))$.

4.3 $2^{Nc(USC(V))} = Nc(V)$ (4.2 and 2.5)

4.4 $2^{Nc(USC^2(V))} = Nc(USC(V))$ (4.2 and 2.7).

4.5 $2^m = \Lambda$ if and only if $Nc(USC(V)) < m$.

4.6 If $2^m \neq \Lambda$, then $m < 2^m$ (R 390).

4.7 $Nc(USC(V)) < Nc(V)$ (4.3 and 4.6).

4.8 If $m \leq n$ and $2^n \neq \Lambda$, then $2^m \neq \Lambda$ and $2^m \leq 2^n$.

4.9 "$2^m = n$" is stratified if "m" and "n" have the same type.

5.1 Definition of $T(m)$ for cardinal numbers m: $T(Nc(a)) = Nc(USC(a))$ (cf. 2.4).

5.2 $T(1) = 1$, $T(2) = 2$, $T(Nc(V)) = Nc(USC(V))$, $T(Nc(USC(V))) = Nc(USC^2(V))$.

5.3 If m, n, $m + n$ are cardinal numbers, then $T(m + n) = T(m) + T(n)$.

5.4 If m is a finite cardinal number, then $m \neq T(m) + 1$ and $m \neq T(m) + 2$ (3.4, 5.2, 5.3).

5.5 If m, n are cardinal numbers, then $m \leq n$ if and only if $T(m) \leq T(n)$ (2.4).

5.6 If $m \leq T(n)$, then there is a p such that $m = T(p)$

5.7 If $m \leq Nc(USC(V))$, then there is a p such that $m = T(p)$.

5.8 $2^{T(m)} \neq \Lambda$ (4.5)

5.9 If $2^m \neq \Lambda$, then $T(2^m) = 2^{T(m)}$ (2.6 and 4.2).

6.1 If m is a cardinal number, then $\Phi(m)$ is the set of cardinal numbers m, 2^m, 2^{2^m}, ... To formalize this, define $Q(m, n)$ for $m, n \in NC$ and $2^m = n$; $\phi(m) = \text{Clos}(\{m\}, Q)$ (R 245; stratification 4.9).

6.2 If $2^m = \Lambda$, then $\Phi(m) = \{m\}$.

6.3 $\phi(Nc(V)) = \{Nc(V)\}$, $Nc(\phi(Nc(V))) = 1$.

6.4 If $n \in \phi(m)$, then $m \leq n$.

6.5 If $2^m \neq \Lambda$, then $m \notin \phi(2^m)$ (4.6).

6.6 If $2^m \neq \Lambda$, then $\phi(m) = \{m\} \cup \Phi(2^m)$.

Proof: (1) $\phi(m) \subseteq \{m\} \cup \phi(2^m)$: $m \in \{m\} \cup \phi(2^m)$; assume $n \in \{m\} \cup \phi(2^m)$ and $2^n \neq \Lambda$; if $n = m$, then $2^n \in \phi(2^m)$; if $n \in \phi(2^m)$, then $2^n \in \phi(2^m)$.

(2) $\phi(2^m) \subseteq \phi(n) - \{m\}$: by 4.6, $2^m \in \phi(m) - \{m\}$; assume $n \in \phi(m) - \{m\}$ and $2^n \neq \Lambda$. So $2^n \in \phi(m)$; by 6.4, $m \leq n$; by 4.6 and 4.8, $m < 2^m \leq 2^n$, so $2^n \in \phi(m) - \{m\}$.

6.7 If $2^m \neq \Lambda$, then $Nc(\phi(m)) = Nc(\phi(2^m)) + 1$ (6.5 and 6.6)

6.8 If $2^m = \Lambda$, then $Nc(\phi(T(m))) = 2$ or 3.

Proof: By the hypothesis and 4.5, $Nc(USC(V)) < m$; so by 5.2, 5.5, $T(m) \geq T(Nc(USC(V))) = Nc(USC^2(V))$. So by 4.8 and 4.4, $2^{T(m)} \geq 2^{Nc(USC^2(V))} = Nc(USC(V))$. If $2^{T(m)} > Nc(USC(V))$, then $\phi(T(m)) = \{T(m), 2^{T(m)}\}$. If $2^{T(m)} = Nc(USC(V))$, then by 4.3 $2^{Nc(USC(V))} = Nc(V)$ and $\phi(T(m)) = \{T(m), Nc(USC(V)), Nc(V)\}$.

7.1 If $\phi(T(m))$ is finite, so is $\phi(m)$.

Proof by induction on $Nc(\phi T(m))$: If $2^m = \Lambda$, then $\phi(m) = \{m\}$. If $2^m \neq \Lambda$, then by 5.8, 5.9, 6.7, $Nc(\phi(T(m))) = Nc(\phi(2^{T(m)})) + 1 = Nc(\phi(T(2^m))) + 1$; so by 3.6, $Nc(\phi(T(2^m))) < Nc(\phi(T(m)))$. If $\phi(2^m)$ is finite, so is $\phi(m)$ by 6.7 and 3.3.

7.2 If $\phi(m)$ is finite, so is $\phi(T(m))$ and $Nc(\phi(T(m))) = T(Nc(\phi(m))) + k$, where $k = 1$ or $k = 2$.

Proof by induction on $Nc(\phi(m))$: (We have achieved stratification by introducing a "T" on the right-hand side.) Assume $Nc(\phi(m)) = 1$; so $\phi(m) = \{m\}$, $2^m = \Lambda$; so by 6.8, $Nc(\phi(T(m))) = 2$ or 3, by 5.2, $2 = T(1) + 1$, $3 = T(1) + 2$. Assume $Nc(\phi(m)) > 1$; so $2^m \neq \Lambda$ and $Nc(\phi(m)) = Nc(\phi(2^m)) + 1$. By 3.6, $Nc(\phi(2^m)) < Nc(\phi(m))$; by 5.2, 5.3, and 6.6, $Nc(\phi(T(m))) = Nc(\phi(2^{T(m)})) + 1 = Nc(\phi(T(2^m))) + 1 = T(Nc(\phi(2^m))) + k + 1 = T(Nc(\phi(2^m)) + 1) + k = T(Nc(\phi(m))) + k$, where $k = 1$ or 2.

7.3 There is a cardinal number m such that $\phi(m)$ is finite and $T(m) = m$.

Proof: Let c be the set of cardinal numbers n such that $\phi(n)$ is finite. By 6.3, c is not the null set. Let m be the smallest cardinal number in c; so $\phi(m)$ is finite. By 7.2, $\phi(T(m))$ is finite, so $m \leq T(m)$. By 5.6, there is a cardinal number p such that $m = T(p)$; $T(p) \leq T(T(p))$ and by 5.5, $p \leq T(p)$. By 7.1, $\phi(p)$ is finite, so $p = T(p)$, $m = T(m)$.

7.4 There is a finite cardinal number n such that $n = T(n) + 1$ or $n = T(n) + 2$.

Proof: Choose m such that $\phi(m)$ is finite and $T(m) = m$ and let $n = Nc(\phi(m))$. By 7.2, $n = Nc(\phi(T(m))) = T(Nc(\phi(m))) + k = T(n) + k$, where $k = 1$ or 2.

7.5 Contradiction: 5.4 and 7.4.[4]

8.1 Generalized continuum hypothesis in "New Foundations": If m, 2^m, n are cardinal numbers, m not finite and $m \leq n \leq 2^m$, then either $m = n$ or $n = 2^m$. The generalized continuum hypothesis does not hold in "NF."[5] The proof is by proving the theorem of Lindenbaum and Tarski in "NF" according to which the axiom of choice is a consequence of the generalized continuum hypothesis.

2 212

[1] Quine, W. V., "New Foundations for Mathematical Logic," *Am. Math. Monthly*, **44**, 70–80 (1937).

[2] The axiom of infinity has been proved in a paper submitted to the *Journal of Symbolic Logic;* the constant use of cardinal number in the present note goes back to the referee of that paper.

[3] Rosser, J. B., *Logic for Mathematicians*, McGraw-Hill Book Company, Inc., New York, 1953. References to this book will be of the form "(R 256)," where the number indicates the page.

[4] By a slight modification, we can prove the following theorem (without the axiom of choice): If $m = Nc(a) = Nc(SC(a))$, then there is a cardinal number n such that neither $n \leq T(m)$ nor $T(m) \leq n$. A finite set is therefore not equivalent to its power set; this has been proved in the paper mentioned in reference 2.

[5] This has been proved in the paper mentioned in reference 2 with the axiom of choice.

NUCLEAR REACTIONS WITH ENERGETIC NITROGEN IONS

By H. L. Reynolds, D. W. Scott, and A. Zucker

Oak Ridge National Laboratory

Communicated by Gregory Breit, June 29, 1953

1. Introduction.—The possibility of obtaining information concerning the nucleus with accelerated heavy ions was discussed recently by Breit, Hull, and Gluckstern.[1] Several sources of energetic heavy nuclei have been reported. Sextuply-charged carbon ions have been accelerated in the 60-inch cyclotron[2] at Berkeley and in the 170-inch synchrocyclotron[3] at the University of Chicago to energies of 120 Mev and 1000 Mev, respectively. The Berkeley group has reported several reactions produced by carbon ions, including some reactions resulting in the production of californium.[4] Recently, Miller[5] investigated nuclear reactions occurring in emulsions exposed to the external carbon beam of the 60-inch cyclotron. Heavy nuclei have been observed in nuclear emulsions exposed to cosmic rays at high altitudes, but, due to the high energies and low fluxes of these cosmic-ray particles,[6] they do not lend themselves conveniently to a study of nuclear properties at low energies.

The experiments reported here utilized triply-charged nitrogen ions accelerated in the ORNL 63-inch cyclotron to an $H\rho$ corresponding to an energy of approximately 25 Mev. They represent an extension of the activation experiments of Wyly and Zucker[7] in which radioactive products from nitrogen bombarded targets are detected and identified mainly from their half-lives. The choice of nitrogen as the accelerated particle was based on the following considerations: (1) the magnetic field available to the resources of the program was limited to an $H\rho$ of 3.75×10^5 oersted-inches; in order that the accelerated particles penetrate the coulomb

barrier of light nuclei the choice of bombarding nucleus was thus limited to elements lighter than neon; (2) lithium, boron, beryllium, and fluorine ions are not readily produced in a compact cyclotron ion source due to their refractory nature or corrosive effects; (3) of the three most likely remaining elements—carbon, nitrogen, and oxygen—the low binding energy of nitrogen makes possible a maximum number of reactions at a given energy; and (4) the unique e/m ratio of triply-charged nitrogen ions eliminates the possibility of another nuclear species being accelerated. An ion source for the production of triply-charged nitrogen ions has been

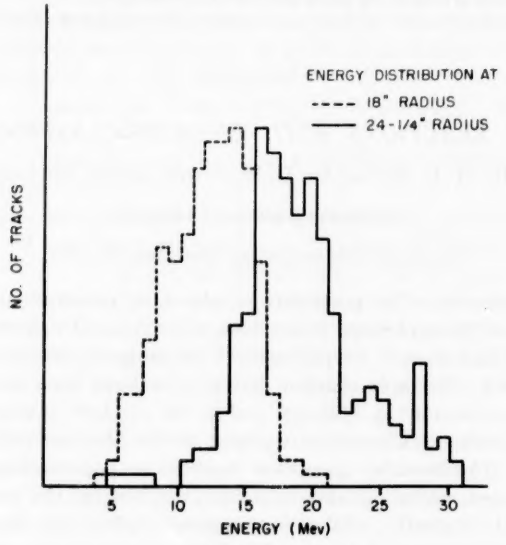


FIGURE 1

Energy distribution of nitrogen ions at radii of 18" and 24¼" as determined from emulsion exposures.

developed by Jones and Zucker.[8] This source also produces a large current of singly-charged nitrogen ions which are resonant at one-third of the oscillator frequency. The singly charged particles do not produce reactions since their maximum energy in this cyclotron is less than 3 Mev.

2. Experimental Method.—Targets were bombarded on a water-cooled probe which was inserted into the cyclotron vacuum tank through an air lock. Unless otherwise specified, all targets were bombarded at a radius of 24.25 inches, where the $H\rho$ for triply-charged ions corresponds to an energy of 25 Mev. The energy spectrum at this radius is shown in figure 1. A beam current of approximately one microampere was held constant

durir.g the long bombardments by adjusting the dee voltage for a constant rate of neutron production as monitored by a BF_3 counter. After bombardments lasting from one to thirty minutes the targets were withdrawn through the air lock and placed under a shielded thin-window Geiger counter. Usually about two minutes elapsed from the time the beam was turned off until counting began. Targets were counted continuously until the counting rate for induced activities approximated background. It was thus possible to detect half-lives ranging from one minute to several days. In general, there were several induced activities in each target. Standard graphical methods were used to determine half-lives and analyze decay curves. From the known bombardment time and rate of decay, the relative thick-target yields for the production of radioactive isotopes were calculated.

The observed radioactive isotopes in this experiment exhibit beta spectra whose maximum energy varies from 0.5 to 4 Mev.[9] It was therefore necessary to correct the data for the relative efficiency of beta counting. At 25 Mev, nitrogen ions have a total range of approximately 6 mg./cm.^{2} The reactions probably occur in the first 2 mg./cm.^{2} of target thickness. The Geiger counter end window was 2 mg./cm.^{2} thick. A correction for β-particle absorption was found to be necessary only in the case of F^{18}, where it amounted to 4%. Multiple scattering in these absorbers has been neglected. Targets were backed by a sufficient thickness of aluminum, while being counted, to assure that source back-scattering did not affect the relative beta-counting efficiency.[10] Scattering from the aluminum lining of the lead shield, from the sample holder, and from the air surrounding the sample was considered negligible.

Whenever possible, pure target elements such as graphite or amorphous boron were first bombarded in order to establish the reactions which occurred in any one element. Where this method was impractical, compounds such as ZnO were used. The heavier elements in these compounds were chosen such that the coulomb barriers were too high to be penetrated by 25-Mev nitrogen nuclei. Compounds of the light elements were used to determine the relative reaction yields of their constituents, for example, boron carbide (B_4C) was used to determine the reaction yields in boron relative to those in carbon. It was not necessary to measure beam currents, nor was the knowledge of relative stopping power of the target constituents required to determine relative yields.

3. Results.—In this section target elements are described in the order in which they were bombarded to obtain a consistent set of relative yields. The results are summarized in table 1. An arbitrary yield of 100 is assigned to the reaction[11]

$$C^{12}(N^{14})Na^{24}2H^1.$$

In the table the bombarded elements are listed in the order of increasing Z and the individual reactions in the order of decreasing Q.[12] All relative yields are corrected for isotopic abundance of the target isotope and, where necessary, for beta-ray absorption.

TABLE 1

NITROGEN REACTIONS IN LIGHT NUCLEI—THICK TARGET RELATIVE YIELDS

REACTION[11]	$T_{1/2}$	$\frac{Y}{YC^{12}(N^{14})Na^{24}2H^1}$	$\frac{Q}{Mev}$	$\frac{E_{COM}}{Mev}$	$\frac{E_{Barr.}}{Mev}$
$D^2(N^{14})O^{15}n^a$	2.1 min.	3600	+ 5.1	2.4	3.3
$Be^9(N^{14})Na^{22}n$	2.6 yr.	<200	+11.2	7.4	5.0
$Be^9(N^{14})F^{18}He^4n$	112 min.	6500	+ 3.0		
$Be^9(N^{14})N^{13}Be^{10}$	10.2 min.	<1.0	− 3.8		
$Be^9(N^{14})F^{17}He^6$	70 sec.	<2.0	− 5.5		
$B^{10}(N^{14})Na^{22}H^2$	2.6 yr.	<400	+ 6.8	7.9	6.4
$B^{10}(N^{14})C^{11}C^{13}$	20.5 min.	30	+ 1.1		
$B^{10}(N^{14})N^{13}B^{11}$	10.2 min.	15	+ 0.90		
$B^{10}(N^{14})O^{15}Be^9$	2.1 min.	60	+ 0.80		
$B^{10}(N^{14})F^{18}Li^6$	112 min.	20	+ 0.20		
$B^{11}(N^{14})Na^{24}H^1$	15.0 hr.	20	+12.7	8.4	6.5
$B^{11}(N^{14})Ne^{23}2H^1$	40.2 sec.	0.5	+ 2.6		
$C^{12}(N^{14})Na^{22}He^4$	2.6 yr.	630	+ 5.4	8.8	7.9
$C^{12}(N^{14})F^{18}Be^8$	112 min.	70	− 2.8		
$C^{12}(N^{14})Na^{24}2H^1$	15.0 hr.	100	− 3.2		
$C^{12}(N^{14})N^{13}C^{13}$	10.2 min.	<0.2	− 5.7		
$C^{13}(N^{14})Na^{25}2H^1$	62 sec.	140	+ 1.0	9.2	7.7
$N^{14}(N^{14})N^{13}N^{15}$	10.2 min.	20	+ 0.23	9.5	8.6
$N^{14}(N^{14})O^{15}C^{13}$	2.1 min.	40	− 0.27		
$O^{16}(N^{14})Al^{26}2H^1$	2.3 min.	250	+ 0.73	10.1	9.8
$O^{16}(N^{14})Na^{22}Be^8$	2.6 yr.	<90	− 1.7		
$O^{16}(N^{14})F^{18}C^{12}$	112 min.	9.0	− 2.6		
$O^{16}(N^{14})P^{30}\gamma$	2.5 min.	<25	+18.4		
$O^{16}(N^{14})O^{15}N^{15}$	2.1 min.	<25	− 4.8		
$O^{16}(N^{14})N^{13}O^{17}$	10.2 min.	<0.05	− 6.4		
$S^{32}(N^{14})Ti^{46}H^1$	3.1 hr.	<30	+10.2	13.2	18.5
$S^{32}(N^{14})Sc^{44}2H^1$	4.0 hr.	300	+ 1.8		
$S^{32}(N^{14})Sc^{44*}2H^1$	2.4 d.	150	+ 1.5		
$S^{32}(N^{14})Cl^{34}C^{12}$	33 min.	10	+ 0.9		
$S^{34}(N^{14})V^{47}n$		250	+ 7.5		
$S^{32}(N^{14})F^{18}Si^{28}$	112 min.	<15	− 2.6		
$S(N^{14})$??	2.7 min.	50	...		

[a] Average absolute cross-section for this reaction for energy spectrum at 24¼-inch radius (average energy 19 Mev) is 3×10^{-27} cm².

The table shows possible reactions with a Q greater than minus 6 Mev, and a half-life longer than one minute. B^{11}, C^{13}, and N^{14} were perforce bombarded only in targets containing other light isotopes Because of the

uncertainties introduced by other elements in the targets, table 1 lists only the observed activities in B^{11}, C^{13}, and N^{14} targets, and does not list other energetically possible reactions. Also, all possible sulfur reactions were not listed because of the large number of possibilities and the uncertainty in product mass values. The center-of-mass energy was calculated for the average of the energy distribution, 19 Mev in the laboratory system. However, the energy spectrum extends up to 30 Mev, as is shown in figure 1. The coulomb barriers were calculated from nuclear radii which were determined from the total fast neutron cross-section given by Coon[13] and the theory of Feshbach and Weisskopf.[14]

A knowledge of half-life is usually sufficient for identification in target elements lighter than fluorine. The heaviest nucleus which can be formed by bombarding oxygen with nitrogen is phosphorus. The number of isotopes with Z less than 15 and with half-lives greater than 30 seconds is well known. There are a few isotopes with half-lives nearly identical, as will be shown in the case of oxygen targets. The probability of producing unreported isotopes with half-lives greater than 30 seconds is small in this part of the periodic table. Light element isotope mass values are well known and the list of probable reactions is narrowed severely by the limited amount of energy available in the COM system. An attempt was not made to determine half-lives accurately. Once the products resulting from a target element were identified, the published half-lives were used in order to deduce the proper relative activities from the decay scheme. In several cases characteristic gamma-ray energies were measured to aid in the identification of reaction products.

A few elements between oxygen and sulfur have been bombarded. The complexity of the decay curves and the many possible products which may be formed indicate that proper identification could not be made without the use of chemical separation.

Carbon: Carbon was bombarded in the form of graphite strips $^1/_{32}$ inch thick. The activities observed after nitrogen bombardment of carbon had half-lives of 112 minutes, 15 hours, and longer than one week. These activities are assigned to F^{18}, Na^{24}, and Na^{22}. Sodium-22 was identified by its characteristic 1.28-Mev gamma ray, which was measured with a NaI scintillation spectrometer.

There are several other energetically possible nitrogen-induced reactions in carbon, such as $C^{12}(N^{14})N^{13}C^{13}$ or $C^{12}(N^{14})C^{11}N^{15}$, leading to half-lives of 10.1 and 20.5 minutes. If the yields of these reactions were small, the half-lives would be masked by the presence of 112-minute F^{18}. Taking advantage of the fact that the maximum F^{18} beta-ray energy is 0.65 Mev, about one-half of the maximum energy of C^{11} or N^{13}, aluminum foils were used to absorb the F^{18} beta-particles. With the absorber present, the only new activity detected had a half-life of approximately one minute. We can

assign an upper limit of 0.1 to the relative yields for the production of N^{13} or C^{11}. The one-minute activity is probably due to Na^{25} produced in the reaction $C^{13}(N^{14})Na^{25}2H^1$.

Deuterium: A zirconium deuteride target was prepared by evaporating a zirconium film on a platinum backing and heating it in an atmosphere of deuterium. After bombardment, a 2-minute activity was observed which

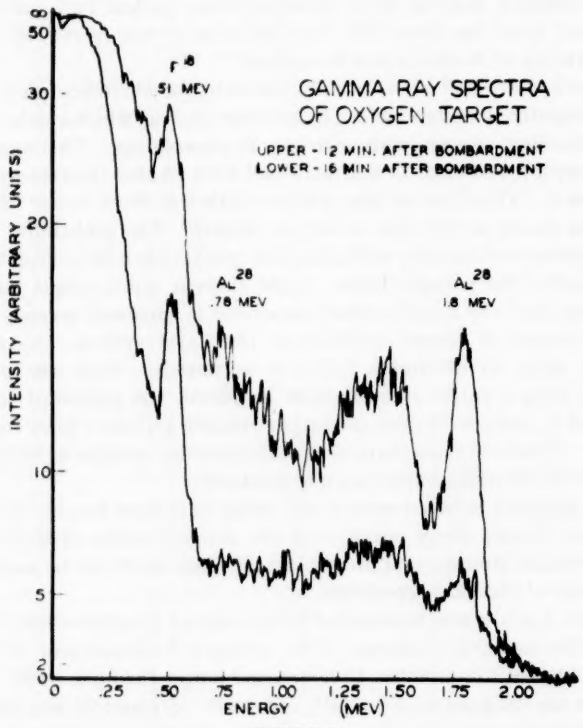


FIGURE 2

Gamma-ray spectra of oxygen target taken 12 and 16 minutes after bombardment to illustrate the decay of Al^{28}.

was assigned to O^{15} from the reaction $H^2(N^{14})O^{15}n$. Di-deutero-cupreine[15], $(C_2D_2)_n$, was bombarded to relate the O^{15} yield from deuterium to the Na^{24} yield from carbon.

Boron: Natural amorphous boron, (19% B^{10}, 81% B^{11}) and separated B^{10}, (96% B^{10}), were used as targets. A thick slurry of boron in water was dried on silver or copper to form the target. The following activities were

observed with natural boron: 15-hour Na^{24}, 112-minute F^{18}, 20-minute C^{11}, 10-minute N^{13}, and 2-minute O^{15}. By bombarding separated B^{10} it was determined that the Na^{24} activity is the result of the reaction B^{11}-$(N^{14})Na^{24}H^{1}$. It is possible that some of the reactions listed as due to B^{10} may in part be due to B^{11}. Table 1 indicates that, in general, high Q reactions are favored. Since the Q for all of the reactions in question is at least 3 Mev higher in B^{10} than in B^{11}, we are inclined to assign the reactions leading to C^{11}, N^{13}, F^{18}, and O^{18} to reactions with B^{10}. The reaction B^{11}-$(N^{14})Ne^{23}2H^{1}$ was apparently being obscured by the activities of the other reaction products. Neon 23 has a half-life of 40 seconds and a beta end-point energy of 4.3 Mev. Since the Ne^{23} maximum beta energy is approximately twice as large as that from any of the other boron reaction products, its presence was observed by using appropriate absorbers during counting.

Boron carbide was used as a target to relate the yields from boron to the Na^{24} yield from the carbon target. The target was prepared in the same manner as the boron targets.

Oxygen: Oxygen was bombarded as TiO_2, ZnO, and PbO slurries dried on silver. Two activities were found: a 112-minute activity assigned to F^{18}, and a short-lived activity of 2.2 minutes. The short activity could be due to any of the following reactions: $O^{16}(N^{14})P^{30}\gamma$, $O^{16}(N^{14})Al^{28}2H^{1}$, or O^{16}-$(N^{14})O^{15}N^{15}$. A gamma ray of 1.8 Mev, presumably due to Al^{28}, was identified with a NaI spectrometer and found to be connected with the short-lived activity. We did not observe annihilation quanta which would result from the 2-minute positrons of O^{15} or P^{30}. We therefore assign the 2.2-minute activity to Al^{28}, with the formation of P^{30} or O^{15} less probable by a factor of ten. Figure 2 shows the gamma spectra of a nitrogen-bombarded oxygen target taken 12 and 16 minutes after bombardment. The 1.8-Mev peak decreases approximately by a factor of four over the 4-minute period. An upper limit for the relative yield of N^{13} was determined by using an absorber to stop the beta particles from F^{18}. The oxygen reactions were related to the carbon standard reaction with $SrCO_3$.

Beryllium: The beryllium target was a 5-mil foil obtained commercially. A very strong activity of 112-minute half life was observed which is assigned to F^{18}. No other activities were observed with an absorber to stop the F^{18} beta particles. A $^{1}/_{16}$-inch thick BeO target was used to tie the Be reaction yield to that of oxygen which was previously related to the carbon yield.

Nitrogen: A satisfactory nitrogen target free from other light elements was not available. Tantalum nitride prepared by passing nitrogen gas over hot tantalum was found to contain carbon and oxygen contamination. Silver cyanide was satisfactory for the detection of nitrogen-nitrogen reaction yields, and at the same time served to relate them to carbon reaction yields. The production of F^{18} or Na^{24} from a nitrogen target was

unlikely because of the very negative Q values and therefore a direct comparison of the nitrogen products with those of carbon could be made. Half-lives of two minutes and ten minutes were produced by nitrogen-nitrogen reactions, and are assigned to O^{15} and N^{13}, respectively.

Sulfur: Sulfur was bombarded as amorphous sulfur powder pressed on a silver backing. Four activities were observed: 2.4-day, 4-hour, 40-minute, and 2.7-minute. The 2.4-day activity may be assigned to an isomeric state of Sc^{44} produced in the reaction $S^{32}(N^{14})Sc^{44*}2H^1$. There are two possible 4-hour activities in this region of the periodic table: the decay of the ground state of Sc^{44} and Sc^{43} which exhibit a 1.16-Mev[16] and a 0.37-Mev[17] gamma ray, respectively A gamma-ray spectrum showed a 1.15-Mev gamma ray associated with the 4-hour activity, whereas a 0.35-Mev gamma ray was not found. We conclude that the greater part of the 4-hour activity is due to the reaction $S^{32}(N^{14})Sc^{44}2H^1$. The observed 40-minute activity may be due to the 33-minute activity from the reactions $S^{32}(N^{14})C^{12}Cl^{34}$ or $S^{34}(N^{14})V^{47}n$. Both possibilities are shown in table 1, the latter corrected for the 4.2% isotopic abundance of S^{34}. The 2.7-minute activity was not assigned and may be a composite of a number of short half-lives. $SrSO_4$ was used to relate the yields of sulfur to those of oxygen.

4. Beam Energy Spectrum.—Since beam orbit centers may wander considerably from the geometric center of the cyclotron, knowledge of the magnetic field and the radial position of the target does not necessarily specify the particle energy. Therefore, an independent method of energy determination was required. Ilford C-2 emulsions were covered with 0.18 mg./cm.^{2} aluminum foils and placed in the beam at the same radius as the targets. A range spectrum was then obtained from the plates. The range-energy relation for nitrogen ions in emulsion has been obtained experimentally by Reynolds and Zucker.[18] These data were used to transform the beam range distribution into the energy distribution shown in figure 1. The energy spectrum is broad, with an average energy of approximately 19 Mev. The greatly reduced beam currents required for emulsion exposure were obtained by reducing either the dee voltage or the ion source current. Reduction of the dee voltage by a factor of five gave an energy spectrum at $24^1/_4$ inches whose peak was at an energy lower by one or two Mev than the peak of the spectrum obtained with reduced ion source current and full dee voltage. It is believed that reducing the arc current causes a negligible change in the spectrum.

The geometry of the emulsion exposures was of necessity somewhat different from that of the target exposure. Nevertheless, a carbon target which was bombarded under the same geometry as the emulsion gave the same relative yield as carbon targets exposed in the usual manner. Since it will be shown that the relative yields from carbon are sensitive to energy,

the energy spectrum shown in figure 1 represents the distribution of particles which reached the activation targets.

5. Energy Dependence.—The energy dependence of the relative yields in carbon was investigated by bombarding targets at various radii. A clipping probe was placed 90 deg. in front of the target at a geometric radius $^1/_4$ inch larger than that of the target. This probe reduced the number of ions reaching the target with energy higher than that calculated from the radius. Figure 1 shows the energy spectrum at 18 inches, which may be compared with the spectrum at 24.25 inches in the same figure. Table 2 shows the yields of the reactions $C^{12}(N^{14})F^{18}Be^8$ and $C^{12}(N^{14})Na^{22}He^4$ relative to the yield of $C^{12}(N^{14})Na^{24}2H^1$, as a function of radius.

TABLE 2

REACTION YIELDS FROM NITROGEN BOMBARDMENT OF CARBON AS A FUNCTION OF RADIUS OF BOMBARDMENT

RADIUS (IN.)	$C^{12}(N^{14})Na^{24}2H^1$	$C^{12}(N^{14})Na^{22}He^4$	$C^{12}(N^{14})F^{18}Be^8$
$24^1/_4$	100[a]	630	70
20	100[a]	3,800	4.3
18	100[a]	35,000	3.8

[a] Assigned value as reference point.

6. Experimental Errors.—This has been primarily a survey experiment and accuracy in the determination of relative yields has not been a prime consideration. Errors could arise from the following causes:

1. Counting errors: these have been considered previously.
2. Target impurities: commercially available target materials were usually pure to a few parts per thousand; in cases where unexpected impurities appeared, in TaN, for example, they were readily detected by the anomalous decay curve.
3. Changes in the target during bombardment: stable compounds with high dissociation temperatures were chosen as target materials in order to avoid changes in the chemical composition of the target during bombardment; targets bombarded with reduced beams gave the same results, within the limits of accuracy, as those bombarded with intense beams.
4. Analysis of decay schemes: decay curves containing more than two or three half-lives are difficult to analyze accurately; with several independent analyzers and a variety of targets, we were able to reproduce relative yield values to within the estimated error.
5. Variation of beam energy distribution: relative yields depend on the beam energy spectrum, *vide supra;* errors due to this cause were investigated by bombarding the same target element a number of times and by bombarding carbon targets periodically. No serious changes in relative yields were found.

We estimate that the above factors contributed a probable error in relative cross sections of less than $\pm 20\%$ for any one target element, and less than $\pm 50\%$ from one element to another.

7. Discussion.—The range of the nitrogen ions has been shown to be a

linear function of the energy in the energy region of these experiments.[18] Hence the yield can be expressed as

$$Y = AN \int_{E_{\min.}}^{E_{\max.}} p(E_i) \int_0^{E_i} \sigma(E) dE \, dE_i$$

where $p(E_i)$ represents the incident energy spectrum normalized so that

$$\int_{E_{\min.}}^{E_{\max.}} p(E_i) dE_i = 1$$

N is the number of target nuclei per cubic centimeter and A is the range energy proportionality constant. The absolute cross-section as a function of energy for the reaction $N^{14}(D^2)O^{15}n$ has been measured by Newson.[19] This reaction is equivalent to the $D^2(N^{14})O^{15}n$ reaction for equal energies in the center-of-mass system. The average absolute cross-section for the reactions produced by the incident energy spectrum, shown in figure 1, is given by

$$\int_{E_{\min.}}^{E_{\max.}} P(E_i) \frac{\int_0^{E_i} \sigma(E) dE \, dE_i}{\int_0^{E_i} dE}$$

For the reaction $D^2(N^{14})O^{15}n$, this cross-section is 3×10^{-27} cm^2. Other reactions listed in table 1 have average cross-sections in the same proportion as their relative yields.

Fusion of a large portion of the nuclear matter takes place in many of the reactions observed. Presumably, these reactions proceed through the formation of a compound nucleus and subsequent evaporation of a few light particles. Further investigation will be necessary to determine if other processes are operative in reactions where the reaction products have masses similar to those of the reacting nuclei.

In these preliminary experiments only a few of the possible target elements have been bombarded. When the cyclotron beam is deflected, better definition of beam energy and current will make more precise measurements possible.

Acknowledgments.—We are indebted to Mr. R. L. Quinn who developed the scintillation counter used in this work, to Mr. H. C. Hoy and the cyclotron staff, and to Miss M. Gillham and Mrs. L. F. Doss who assisted with the counting. We gratefully acknowledge our discussions with Prof. G. Breit, Dr. R. S. Livingston, and Dr. J. L. Fowler. Professor L. D. Wyly participated in the early stages of this research.

[1] Breit, G., Hull, M. H., Jr., and Gluckstern, R. L., *Phys. Rev.*, **87,** 74 (1952).

[2] Alvarez, L. W., *Ibid.*, **58,** 192 (A) (1940); Condit, R., *Ibid.*, **62,** 301 (A) (1942); York, H., Hildebrand, R., Putnam, T., and Hamilton, J. G., *Ibid.*, **70,** 446 (A) (1946); Miller, J. F., *Ibid.*, **83,** 1261 (L) (1951); Miller, J. F., Hamilton, J. G., Putnam, T. M., Raymond, H. R., and Rossi, G. B., *Ibid.*, **80,** 486 (L) (1950).

[3] Chou, C. N., Fry, W. F., and Lord, J. J., *Ibid.*, **87**, 671 (L) (1952)

[4] Ghiorso, A., Thompson, S. G., Street, K., and Seaborg, G. T., *Ibid.*, **81**, 154 (L) (1951); Rasmussen, J. O., Thompson, S. G., and Ghiorso, A., *Ibid.*, **89**, 33 (1953).

[5] Miller, J. F., University of California Radiation Laboratory Report 1902 (unpublished).

[6] Kaplon, M. F., Peters, B., Reynolds, H. L., and Ritson, D. M., *Phys. Rev.*, **85**, 295 (1952).

[7] Wyly, L. D., and Zucker, A., *Ibid.*, **89**, 524 (1953).

[8] Jones, R. J., and Zucker, A., *Rev. Sci. Inst.*, (in press).

[9] Unless otherwise stated, the National Bureau of Standards "Nuclear Data" compilation has been used for values of half-life and beta end-point energies.

[10] Burtt, B. P., *Nucleonics*, **5**, 28, August, 1949.

[11] The usual reaction notation has been changed slightly. The first element represents the target nucleus with the bombarding particle enclosed in brackets. All reaction products are listed on the right side of the brackets since, in many reactions, the products have nearly equal charge and mass and it would be meaningless to distinguish between them.

[12] Mass values were obtained from Hornyak, W. F., Lauritsen, T., Morrison, P., and Fowler, W. A., *Rev. Mod. Phys.*, **22**, 291 (1950) or Metropolis, W., and Reitwiesner, G., "Table of Atomic Masses," USAEC Report NP-1980 (1950).

[13] Coon, J. H., Graves, E. R., and Borschall, H. H., *Phys. Rev.*, **88**, 562 (1952).

[14] Feshbach, H., and Weisskopf, V. F., *Ibid.*, **76**, 1550 (1949).

[15] We wish to thank Dr. P. S. Rudolph of the Oak Ridge National Laboratory Chemistry Division who prepared this compound.

[16] Bruner, J. A., and Langer, L. M., *Phys. Rev.*, **79**, 606 (1950).

[17] Hoskings, J. R., Duval, J. E., Cheng, L. S., and Kurbatov, J. D., *Ibid.*, **88**, 876 (1952).

[18] Reynolds, H. L., and Zucker, A., *Ibid.*, **91**, 496 (1953).

[19] Newson, H., *Phys. Rev.*, **51**, 620 (1937).

THE DIFFERENTIATION OF GASTRULA ECTODERM IN MEDIUM CONDITIONED BY AXIAL MESODERM

BY M. C. NIU AND V. C. TWITTY

STANFORD UNIVERSITY

Communicated June 23, 1953

The widely accepted conclusion that embryonic "inductors" can transmit their stimuli only when in direct physical contact with the reacting cells has seemed to rest on the best of evidence. Separation of the two tissues by the smallest gap, or interposition of the thinnest cellular or artificial barriers, have invariably proved to block inductive action. In fact, so consistent have been the findings in this respect that it has not been warranted, in spite of the great amount of work on the chemistry of inductors, to assert positively that diffusible substances are involved in the inductive control of differentiation. The results of Holtfreter[1] and others on the neuralization of young gastrula ectoderm in solutions of various sub-

stances do not constitute true exceptions, since according to Holtfreter the action of these agents was probably indirect, through toxic damage which activates strictly endogenous mechanisms.

The present study was originally undertaken with the object of investigating inductive relationships in very small cell populations. Encouraged by earlier success in the cultivation of propigment cells isolated singly or in groups of two or more,[2] it was hoped that it might be possible to work on a similar scale with dissociated chordamesoderm and gastrula ectoderm cells, perhaps bringing together representatives of each in small groups or even single pairs to test for any demonstrable effects upon their differentiation. At first technical difficulties, and later other considerations, led to the gradual modification of this approach, and the procedure eventually adopted was to isolate minute pieces of young gastrula ectoderm, the smallest that can readily be excised by ordinary microsurgical methods, together with pieces of axial mesoderm of more substantial size. The mesodermal explants consisted variously of: dorsal blastoporal lip; posterior medullary plate (the portion which invaginates during late gastrulation and forms the mesodermal structures of the tail and posterior part of the trunk); and blocks of mesodermal somites taken from embryos in young tailbud stages of development. In preliminary trials the mesodermal and ectodermal explants were placed in close apposition and often underwent intimate fusion, but it is important to emphasize that in this paper we are concerned only with experiments in which the two tissues were separately situated and had no physical contact with one another.

The species employed were *Triturus torosus*, *T. rivularis*, and *T. granulosus;* and *Amblystoma tigrinum* and *A. mexicanum*. The culture medium

FIGURE 1

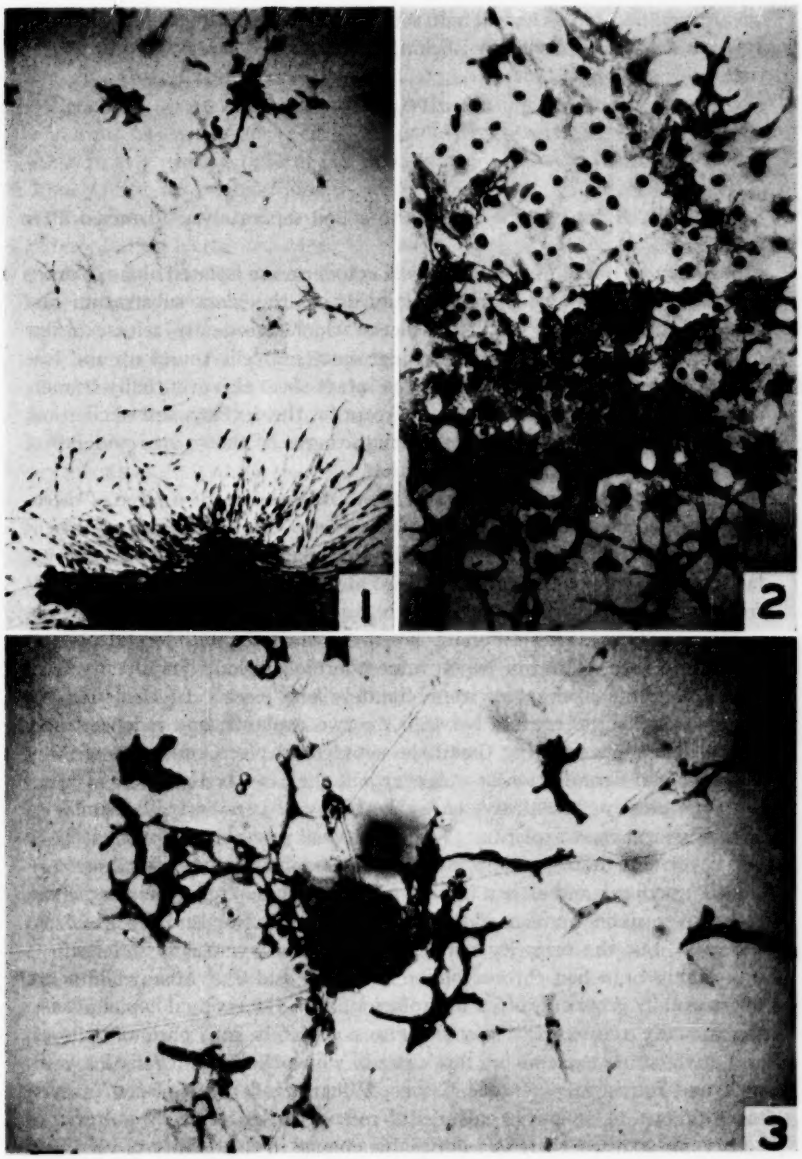
Below, explant of *T. torosus* posterior medullary plate, with outgrowth of differentiating myoblasts. *Above*, scattered chromatophores which have originated from a small piece of torosus gastrula ectoderm (Harrison, stage 10) introduced into the culture 10 days after isolation of the posterior medullary plate. Nerve fibers have grown out of the residual portion of the ectodermal isolate. Photographed 23 days after isolation of the ectoderm.

FIGURE 2

Chromatophores formed from a piece of black axolotl gastrula ectoderm (stage 10) after introduction into a 10-day old culture of *T. torosus* posterior medullary plate. The two explants were situated remotely from one another in the drop. Photographed 12 days after introduction of the ectodermal isolate.

FIGURE 3

Chromatophores and neural tissue (note nerve fibers) formed from a piece of young *T. torosus* gastrula ectoderm isolated in cell-free medium withdrawn from a 6-day-old culture of torosus posterior medullary plate. Photographed 20 days after isolation.


1
2
3

was a modified physiological salt solution devised by us some years ago. Its formula, first published by Flickinger,[3] is repeated here:

SOLUTION A (500cm.³)		SOLUTION B (250cm.³)		SOLUTION C (250cm.³)	
NACl	3400 mg.	Na_2HPO_4	110 mg.	$NaHCO_3$	200 mg.
KCl	50	KH_2PO_4	20		
$Ca(NO_3)\cdot 4H_2O$	80				
$MgSO_4$	100				

Solutions A, B, and C are brought to a boil separately, and mixed after cooling.

When small pieces of young gastrula ectoderm are isolated alone in drops of this medium, they attach temporarily to the glass substratum and spread into very thin epithelial sheets which commonly release a few migrating ameboid cells. Later these scattered cells round up and lose their attachment to the glass, and the intact sheet also eventually retracts and falls free from the cover slip. Except for the development of ciliation, in no case have the explant or the emigrating cells shown any evidence of histological or cytological differentiation.

The ectodermal explant likewise fails to differentiate if a piece of mesoderm has been introduced simultaneously into the same drop. There is active outgrowth from the mesodermal explant, notably of myoblasts, but the behavior of the ectodermal isolate is essentially indistinguishable from that when it is cultivated alone.

The results are very different, however, when the ectodermal piece is introduced into a drop in which a mesodermal explant has already been developing for a period of approximately one week. In the complete absence of physical contact between the two explants, and independently of the distance separating them, the ectodermal piece undergoes striking histological differentiation in well over half the cases tested.

In the most representative cases the behavior is remarkably similar to that of neural crest explants. The ectodermal pieces attach intimately to the cover slip, but spread somewhat less extensively than in "unconditioned" medium, and after a few days begin to give outgrowths of ameboid cells. A variable number of these may later round up and fall free from the glass, but the majority remain attached and eventually differentiate into highly branched chromatophores (Figs. 1 and 2). Meanwhile nerve fibers usually grow out, often in profusion, from the residual explant mass. Occasionally the explant may disperse completely into chromatophores, and at the other extreme are rare cases in which the explant remains compact and forms only nervous tissue. When ectoderm is placed in even older (about 14–18 days) cultures of mesoderm, particularly cultures of embryonic somites, there is a noticeable change in the quality of its differentiation. In general, fewer pigment cells and more nervous tissue are formed, and frequently clearly identifiable myoblasts emerge.

It appears highly improbable that the differentiation of the explants is attributable merely to nonspecific toxic effects of the medium. The explants appear to be completely healthy, and cell dissociation or other evidences of injury are definitely rarer than in unconditioned medium, in which no differentiation occurs. It is also significant that no differentiation ensues when ectoderm pieces are introduced into cultures of endoderm, or into drops containing massive pieces of young epidermis.

The physical presence of the mesodermal explant is not essential to the differentiation of the ectodermal pieces. This has been demonstrated repeatedly by cases in which explants have differentiated into chromatophores and nervous tissue after isolation in cell-free medium drawn from established mesodermal cultures (Fig. 3). In a few series, after the medium had been withdrawn from the latter cultures, it was replaced by fresh physiological salt solution and the mesoderm allowed to continue its differentiation for another week or ten days. When this medium was in turn withdrawn and tested, it often induced ectodermal pieces to give rise to myoblasts in addition to nervous tissue and chromatophores.

The results of the investigation thus seem to indicate clearly that inductor tissues indeed contain, and under the conditions of these experiments can release, diffusible substances capable of effecting "at a distance" the inductive control of ectodermal differentiation; they suggest further that the inductive specificity of the substances released changes with the age or degree of differentiation of the inductor.

The modified physiological salt solution employed for this study has the incidental property of causing ectoderm to attach intimately and spread extensively upon a glass substratum, and it is believed that this fact may have contributed importantly to the positive outcome of the tests. Pieces of ectoderm isolated in Holtfreter's standard salt solution characteristically round up into unattached vesicles, which by virtue of the intact "surface coat" investing them are undoubtedly impervious to larger molecules. This insulating coat is probably torn and disrupted, under the conditions of the present experiments, by the tensions and cellular rearrangements incidental to spreading of the explants into thin sheets, with the result that the constituent cells are partly denuded and thus more accessible to diffusible substances in the medium.

Samples of conditioned medium examined by Dr. Hubert S. Loring of the Department of Chemistry show appreciable light absorption in the ultraviolet with maxima and minima at 265 and 245 respectively, suggestive of nucleic acid or certain of its components. Various experiments are in progress to determine the nature of the substance or substances present.

[1] Holtfreter, J., *J. Exp. Zool.*, **98**, 161–209 (1945).
[2] Twitty, V. C., *Science*, **113**, 476 (1951).
[3] Flickinger, R. A. Jr., *J. Exp. Zool.*, **112**, 465–484 (1949).

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consecutively, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of Journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph.*, Boston, 22, 731–815 (1911); or, Wheeler, W. M., *Konigsberg, Schr. physik. Ges.*, 25, 1–142 (1914).

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, *2101 Constitution Ave., Washington 25, D. C. Articles of which a member of the Academy is author or a co-author may extend to twelve printed pages, those by non-members only to six printed pages; space above these limits will be billed to the author at $17.50 per page.*

PROOF will ordinarily be sent,—with the understanding that charges for an author's correction shall be billed to him. Authors are therefore requested to make final revisions on the typewritten manuscripts.

REPRINTS should be ordered at the time of submission of manuscript. They will be furnished to authors at cost, approximately as follows.

SCHEDULE OF RATES FOR REPRINTS, CARRIAGE CHARGES EXTRA

No. Copies	50	100	150	200	250
4 pp.	4.90	5.60	6.30	7.00	7.70
8 pp.	9.80	11.20	12.60	14.00	15.40
Covers	3.75	4.80	5.85	6.90	7.95

CONTENTS

Page

BIOCHEMISTRY.—ALTERATION OF THE KINETIC PROPERTIES OF AN ENZYME BY THE BINDING OF BUFFER, INHIBITOR, OR SUBSTRATE. . . . . *By Robert A. Alberty and Robert M. Bock* 895

BIOCHEMISTRY.—A LIGHT ACTIVATION PHENOMENON IN THE ENZYMATIC AND NONENZYMATIC REDUCTION OF TETRAZOLIUM SALTS. . . . . *By Walter J. Nickerson and Joseph R. Merhel* 901

BOTANY.—THE INTERACTION OF AUXIN AND LIGHT IN THE GROWTH RESPONSES OF PLANTS. . . . . *By James L. Liverman and James Bonner* 905

CHEMISTRY.—THE CONDUCTANCE OF SOLUTIONS OF SALTS IN ORTHODICHLOROBENZENE . . . . . *By Filippo Accascina, E. Lowell Swarts, Philip L. Mercier, and Charles A. Kraus* 917

GENETICS.—THE EFFECT OF INDOLE-3-ACETIC ACID ON DATURA EMBRYOS . . . . . *By J. Rietsema, S. Satina, and A. F. Blakeslee* 924

GEOPHYSICS.—ON THE ORIGIN OF CONTINENTS AND MOUNTAINS . . . . . *By Harold C. Urey* 933

MATHEMATICS.—BOTTLENECK PROBLEMS AND DYNAMIC PROGRAMMING . . . . . *By Richard Bellman* 947

MATHEMATICS.—ON STEENROD'S REDUCED POWERS, THE INDEX OF INERTIA, AND the TODD GENUS . . . . . *By Friedrich Hirzebruch* 951

MATHEMATICS.—THE IDEAS OF VARIABLE AND FUNCTION . . . *By Karl Menger* 956

MATHEMATICS.—PAIRS OF MATRICES WITH PROPERTY *L* . . . . . *By T. S. Motzkin and Olga Taussky* 961

MATHEMATICS.—ON RESTRICTED PARTITIONS AND A GENERALIZATION OF THE EULER φ NUMBER AND THE MOEBIUS FUNCTION. . . *By Charles A. Nicol* 963

MATHEMATICS.—CHARACTERIZATION OF QUADRATIC FORMS OVER LOCAL FIELDS . . . . . *By O. T. O'Meara* 969

MATHEMATICS.—THE AXIOM OF CHOICE IN QUINE'S NEW FOUNDATIONS FOR MATHEMATICAL LOGIC . . . . . *By Ernst P. Specker* 972

PHYSICS.—NUCLEAR REACTIONS WITH ENERGETIC NITROGEN IONS . . . . . *By H. L. Reynolds, D. W. Scott, and A. Zucker* 975

ZOOLOGY.—THE DIFFERENTIATION OF GASTRULA ECTODERM IN MEDIUM CONDITIONED BY AXIAL MESODERM . . . . *By M. C. Niu and V. C. Twitty* 985